(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number
WO 02/064741 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/04134

(22) International Filing Date: 12 February 2002 (12.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/268,289 13 February 2001 (13.02.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/268,289 (CIP)
Filed on 13 February 2001 (13.02.2001)

(71) Applicant *(for all designated States except US)*: **DIADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SALCEDA, Susana** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **MACINA, Roberto, A.** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **HU, Ping** [US/US]; 108 Palmer Street, San Ramon, CA 94583 (US). **RECIPON, Herve** [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). **KARRA, Kalpana** [IN/US]; 6511 Rainbow Drive, San Jose, CA 95129 (US). **CAFFERKEY, Robert** [IE/US]; 849 West Orange Avenue, Apartment 1030, South San Francisco, CA 94081 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US). **SUN, Yongming** [US/US]; 869 S. Winchester Boulevard, Apartment 260, San Jose, CA 95128 (US).

(74) Agents: **LICATA, Jane, Massey** et al.; Licata & Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/064741 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic breast cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention, The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast tissue, identifying breast tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered breast tissue for treatment and research.

18

M. 60-436

COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application Serial No. 60/268,289 filed February 13, 2001, which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic breast cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast tissue, identifying breast tissue and monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered breast tissue for treatment and research.

BACKGROUND OF THE INVENTION

Excluding skin cancer, breast cancer, also called mammary tumor, is the most common cancer among women, accounting for a third of the cancers diagnosed in the United States. One in nine women will develop breast cancer in her lifetime and about 192,000 new cases of breast cancer are diagnosed annually with about 42,000 deaths. Bevers, *Primary Prevention of Breast Cancer*, in BREAST CANCER, 20-54 (Kelly K Hunt et al., ed., 2001); Kochanek et al., 49 Nat'l.Vital Statistics Reports 1, 14 (2001).

In the treatment of breast cancer, there is considerable emphasis on detection and risk assessment because early and accurate staging of breast cancer has a significant impact on survival. For example, breast cancer detected at an early stage (stage T0, discussed below) has a five-year survival rate of 92%. Conversely, if the cancer is not detected until a late stage (i.e., stage T4), the five-year survival rate is reduced to 13%.

AJCC Cancer Staging Handbook pp. 164-65 (Irvin D. Fleming et al. eds., 5th ed. 1998). Some detection techniques, such as mammography and biopsy, involve increased discomfort, expense, and/or radiation, and are only prescribed only to patients with an increased risk of breast cancer.

Current methods for predicting or detecting breast cancer risk are not optimal. One method for predicting the relative risk of breast cancer is by examining a patient's risk factors and pursuing aggressive diagnostic and treatment regiments for high risk patients. A patient's risk of breast cancer has been positively associated with increasing age, nulliparity, family history of breast cancer, personal history of breast cancer, early menarche, late menopause, late age of first full term pregnancy, prior proliferative breast disease, irradiation of the breast at an early age and a personal history of malignancy. Lifestyle factors such as fat consumption, alcohol consumption, education, and socioeconomic status have also been associated with an increased incidence of breast cancer although a direct cause and effect relationship has not been established. While these risk factors are statistically significant, their weak association with breast cancer limited their usefulness. Most women who develop breast cancer have none of the risk factors listed above, other than the risk that comes with growing older. NIH Publication No. 00-1556 (2000).

Current screening methods for detecting cancer, such as breast self exam, ultrasound, and mammography have drawbacks that reduce their effectiveness or prevent their widespread adoption. Breast self exams, while useful, are unreliable for the detection of breast cancer in the initial stages where the tumor is small and difficult to detect by palpitation. Ultrasound measurements require skilled operators at an increased expense. Mammography, while sensitive, is subject to over diagnosis in the detection of lesions that have questionable malignant potential. There is also the fear of the radiation used in mammography because prior chest radiation is a factor associated with an increase incidence of breast cancer.

At this time, there are no adequate methods of breast cancer prevention. The current methods of breast cancer prevention involve prophylactic mastectomy (mastectomy performed before cancer diagnosis) and chemoprevention (chemotherapy before cancer diagnosis) which are drastic measures that limit their adoption even among women with increased risk of breast cancer. Bevers, *supra*.

A number of genetic markers have been associated with breast cancer. Examples of these markers include carcinoembryonic antigen (CEA) (Mughal et al., 249 JAMA 1881 (1983)) MUC-1 (Frische and Liu, 22 J. Clin. Ligand 320 (2000)), HER-2/neu (Haris et al., 15 Proc.Am.Soc.Clin.Oncology. A96 (1996)), uPA, PAI-1, LPA, LPC, RAK and BRCA (Esteva and Fritsche, *Serum and Tissue Markers for Breast Cancer*, in BREAST CANCER, 286-308 (2001)). These markers have problems with limited sensitivity, low correlation, and false negatives which limit their use for initial diagnosis. For example, while the BRCA1 gene mutation is useful as an indicator of an increased risk for breast cancer, it has limited use in cancer diagnosis because only 6.2 % of breast cancers are BRCA1 positive. Malone et al., 279 JAMA 922 (1998). *See also*, Mewman et al., 279 JAMA 915 (1998) (correlation of only 3.3%).

Breast cancers are diagnosed into the appropriate stage categories recognizing that different treatments are more effective for different stages of cancer. Stage TX indicates that primary tumor cannot be assessed (i.e., tumor was removed or breast tissue was removed). Stage T0 is characterized by abnormalities such as hyperplasia but with no evidence of primary tumor. Stage Tis is characterized by carcinoma in situ, intraductal carcinoma, lobular carcinoma in situ, or Paget's disease of the nipple with no tumor. Stage T1 is characterized as having a tumor of 2 cm or less in the greatest dimension. Within stage T1, Tmic indicates microinvasion of 0.1 cm or less, T1a indicates a tumor of between 0.1 to 0.5 cm, T1b indicates a tumor of between 0.5 to 1 cm, and T1c indicates tumors of between 1 cm to 2 cm. Stage T2 is characterized by tumors from 2 cm to 5 cm in the greatest dimension. Tumors greater than 5 cm in size are classified as stage T4. Within stage T4, T4a indicates extension of the tumor to the chess wall, T4b indicates edema or ulceration of the skin of the breast or satellite skin nodules confined to the same breast, T4c indicates a combination of T4a and T4b, and T4d indicates inflammatory carcinoma. AJCC Cancer Staging Handbook pp. 159-70 (Irvin D. Fleming et al. eds., 5th ed. 1998). In addition to standard staging, breast tumors may be classified according to their estrogen receptor and progesterone receptor protein status. Fisher et al., 7 Breast Cancer Research and Treatment 147 (1986). Additional pathological status, such as HER2/neu status may also be useful. Thor et al., 90 J.Nat'l.Cancer Inst. 1346 (1998); Paik et al., 90 J.Nat'l.Cancer Inst. 1361 (1998);

Hutchins et al., 17 Proc.Am.Soc.Clin.Oncology A2 (1998).; and Simpson et al., 18 J.Clin.Oncology 2059 (2000).

In addition to the staging of the primary tumor, breast cancer metastases to regional lymph nodes may be staged. Stage NX indicates that the lymph nodes cannot be assessed (e.g., previously removed). Stage N0 indicates no regional lymph node metastasis. Stage N1 indicates metastasis to movable ipsilateral axillary lymph nodes. Stage N2 indicates metastasis to ipsilateral axillary lymph nodes fixed to one another or to other structures. Stage N3 indicates metastasis to ipsilateral internal mammary lymph nodes. Id.

Stage determination has potential prognostic value and provides criteria for designing optimal therapy. Simpson et al., 18 J. Clin. Oncology 2059 (2000). Generally, pathological staging of breast cancer is preferable to clinical staging because the former gives a more accurate prognosis. However, clinical staging would be preferred if it were as accurate as pathological staging because it does not depend on an invasive procedure to obtain tissue for pathological evaluation. Staging of breast cancer would be improved by detecting new markers in cells, tissues, or bodily fluids which could differentiate between different stages of invasion. Progress in this field will allow more rapid and reliable method for treating breast cancer patients.

Treatment of breast cancer is generally decided after an accurate staging of the primary tumor. Primary treatment options include breast conserving therapy (lumpectomy, breast irradiation, and surgical staging of the axilla), and modified radical mastectomy. Additional treatments include chemotherapy, regional irradiation, and, in extreme cases, terminating estrogen production by ovarian ablation.

Until recently, the customary treatment for all breast cancer was mastectomy. Fonseca et al., 127 Annals of Internal Medicine 1013 (1997). However, recent data indicate that less radical procedures may be equally effective, in terms of survival, for early stage breast cancer. Fisher et al., 16 J. of Clinical Oncology 441 (1998). The treatment options for a patient with early stage breast cancer (i.e., stage Tis) may be breast-sparing surgery followed by localized radiation therapy at the breast. Alternatively, mastectomy optionally coupled with radiation or breast reconstruction may be employed. These treatment methods are equally effective in the early stages of breast cancer.

Patients with stage I and stage II breast cancer require surgery with chemotherapy and/or hormonal therapy. Surgery is of limited use in Stage III and stage IV patients. Thus, these patients are better candidates for chemotherapy and radiation therapy with surgery limited to biopsy to permit initial staging or subsequent restaging because cancer is rarely curative at this stage of the disease. AJCC Cancer Staging Handbook 84, ¶. 164-65 (Irvin D. Fleming et al. eds., 5th ed. 1998).

In an effort to provide more treatment options to patients, efforts are underway to define an earlier stage of breast cancer with low recurrence which could be treated with lumpectomy without postoperative radiation treatment. While a number of attempts have been made to classify early stage breast cancer, no consensus recommendation on postoperative radiation treatment has been obtained from these studies. Page et al., 75 Cancer 1219 (1995); Fisher et al., 75 Cancer 1223 (1995); Silverstein et al., 77 Cancer 2267 (1996).

As discussed above, each of the methods for diagnosing and staging breast cancer is limited by the technology employed. Accordingly, there is need for sensitive molecular and cellular markers for the detection of breast cancer. There is a need for molecular markers for the accurate staging, including clinical and pathological staging, of breast cancers to optimize treatment methods. Finally, there is a need for sensitive molecular and cellular markers to monitor the progress of cancer treatments, including markers that can detect recurrence of breast cancers following remission.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill in the art from the following description. It should be understood, however, that the following description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat breast cancer and

non-cancerous disease states in breast; identify and monitor breast tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered breast tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to breast cells and/or breast tissue. These breast specific nucleic acids (BSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the BSNA is genomic DNA, then the BSNA is a breast specific gene (BSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to breast. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 66 through 110. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 65. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding a BSP, or that selectively hybridize or exhibit substantial sequence similarity to a BSNA, as well as allelic variants of a nucleic acid molecule encoding a BSP, and allelic variants of a BSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes a BSP or that comprises a part of a nucleic acid sequence of a BSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a BSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a BSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of a BSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of a BSNA.

Another object of the invention is to provided methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is a BSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of a BSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring breast tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered breast tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat breast cancer and non-cancerous disease states in breast. The invention provides methods of using the polypeptides of the invention to identify and/or monitor breast tissue, and to produce engineered breast tissue.

The agonists and antagonists of the instant invention may be used to treat breast cancer and non-cancerous disease states in breast and to produce engineered breast tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the

computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid

sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus

to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such

oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phoshoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its

complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. *See* Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5°C + 16.6 (\log_{10}[Na^+]) + 0.41 (\text{fraction G + C}) - 0.63 (\% \text{ formamide}) - (600/l)$$

where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula: $T_m = 79.8°C + 18.5 (\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.35 (% formamide) - (820/l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula: $T_m = 79.8°C + 18.5(\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.50 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents

to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula: $T_m = 81.5°C + 16.6(\log_{10}[Na^+]) + 0.41(\text{fraction G+C}) - (600/N)$, wherein N is change length and the $[Na^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μg of plasmid or DNA fragment

is digested with about 2 units of enzyme in about 20 µl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAS. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic

acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding a BSP or is a BSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA

molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g.*, Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector.

However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a BSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a

cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation,

GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide

analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$--, and --CH_2SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al., Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least

50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See* Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α-, α-disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, ε-N,N,N-trimethyllysine, ε-N-acetyllysine, O-phosphoserine, N-acetylserine,

N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

1) Serine (S), Threonine (T);
2) Aspartic Acid (D), Glutamic Acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.*, Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default
Max. alignments:	100 (default)
Word size:	11 (default)

No. of descriptions: 100 (default)

Penalty Matrix: BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', $F(ab')_2$, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an $F(ab')_2$ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g.*, Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular

species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g.,* Bird *et al., Science* 242: 423-426 (1988); Huston *et al., Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g.,* Holliger *et al., Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al., Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than1 μM, preferably less than100 nM and most preferably less than 10 nM.

The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "breast specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the breast as compared to other tissues in the body. In a preferred embodiment, a "breast specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "breast specific" nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the breast or to breast cells or tissue or that are derived from such nucleic acid

molecules. These isolated breast specific nucleic acids (BSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to breast, a breast-specific polypeptide (BSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 66 through 110. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 65.

A BSNA may be derived from a human or from another animal. In a preferred embodiment, the BSNA is derived from a human or other mammal. In a more preferred embodiment, the BSNA is derived from a human or other primate. In an even more preferred embodiment, the BSNA is derived from a human.

By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding a BSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a BSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a BSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 66 through 110. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 65.

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under moderate stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 66 through 110. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high

stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 65. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding a BSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human BSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 66 through 110. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a BSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 66 through 110, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a BSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a BSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a BSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 65. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a BSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 65, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a BSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid

molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a BSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to a BSNA or to a nucleic acid molecule encoding a BSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the BSNA or the nucleic acid molecule encoding a BSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 66 through 110 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 65. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the BSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a BSNA. Further, the substantially similar nucleic acid molecule may or may not be a BSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a BSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of a BSNA or a nucleic acid encoding a BSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs

have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a BSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a BSP comprising an amino acid sequence of SEQ ID NO: 66 through 110. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a BSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 65. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a BSP. However, in a preferred embodiment, the part encodes a BSP. In one aspect, the invention comprises a part of a BSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a BSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of a BSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a BSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with ^{33}P, ^{32}P, and ^{35}S, such as α-^{32}P-dATP, α-^{32}P-dCTP, α-^{32}P-dGTP, α-^{32}P-dTTP, α-^{32}P-3'dATP, α-^{32}P-ATP, α-^{32}P-CTP, α-^{32}P-GTP, α-^{32}P-UTP, α-^{35}S-dATP, α-^{35}S-GTP, α-^{33}P-dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY®

TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. *See* Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see* Alers *et al.*, *Genes,*

Chromosomes & Cancer 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.*, Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See* Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. *See* Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene

phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced

with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The Tm of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the Tm of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the Tm by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the Tm by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.*, Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have

discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a BSNA, such as deletions, insertions, translocations, and duplications of the BSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated

herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify BSNA in, and isolate BSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A^+-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to BSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the

probe or primer is derived from a nucleic acid molecule encoding a BSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 66 through 110. In another preferred embodiment, the probe or primer is derived from a BSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 65.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate,

nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998);

Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λGT10 and λGT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as ura3-52, his3-D1, leu2-D1, trp1-D1 and lys2-201.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda* , *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway.

Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription

include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast α-mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the BSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the

polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β-globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PltetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverley, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent

proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α-agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present

invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g,* Cormack *et al.*, *Gene* 173: 33–38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g,*. Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.*, Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See* also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and

NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide BSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and

racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database http://www.abrf.org/ABRF/Research Committees/deltamass/deltamass.html (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. Nucleic Acids Res. 29; 332-335 (2001) and http://www.glycosuite.com/ (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. Nucleic Acids Research, 27: 370-372 (1999) and http://www.cbs.dtu.dk/databases/OGLYCBASE/ (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. Nucleic Acids Res 27(1):237-239 (1999) and http://www.cbs.dtu.dk/databases/PhosphoBase/ (accessed October 19, 2001); or http://pir.georgetown.edu/pirwww/search/textresid.html (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated

compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational

modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g, p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally

modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are

well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae, Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (*e.g.*, from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from breast are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human breast cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further

details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. *See, e.g.*, Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.*, Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.*, with $CaCl_2$, or a solution of Mg^{2+}, Mn^{2+}, Ca^{2+}, Rb^{+} or K^{+}, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (*e.g.*, Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5α competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols (BioRad, Richmond, CA, USA) (http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol

(PEG) and Ca^{2+}. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with $CaPO_4$ or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for $CaPO_4$ transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al.*, *Proc.*

Natl. Acad. Sci. USA 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a breast specific polypeptide (BSP). In an even more preferred embodiment, the polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 66 through 110. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of a BSP. In a more preferred embodiment, the fragment is

derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 66 through 110. A polypeptide that comprises only a fragment of an entire BSP may or may not be a polypeptide that is also a BSP. For instance, a full-length polypeptide may be breast-specific, while a fragment thereof may be found in other tissues as well as in breast. A polypeptide that is not a BSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-BSP antibodies. However, in a preferred embodiment, the part or fragment is a BSP. Methods of determining whether a polypeptide is a BSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g.*, Lerner, Nature 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, a BSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment of polypeptide of the invention, preferably a BSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably a BSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be breast-specific. In a preferred embodiment, the mutein is breast-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 66 through 110. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity,

even more preferably at least 70%, yet more preferably at least 80% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 66 through 110. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 66 through 110.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is breast-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to a BSP. In an even more preferred embodiment, the polypeptide is homologous to a BSP selected from the group having an amino acid

sequence of SEQ ID NO: 66 through 110. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a BSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 66 through 110. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 66 through 110. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 66 through 110. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 66 through 110. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a BSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a BSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the BSNA is selected from the group consisting of SEQ ID NO: 1 through 65. In another preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a BSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the BSP is selected from the group consisting of SEQ ID NO: 66 through 110.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 66 through 110. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the BSP is a member of a family of

polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a BSP. Further, the homologous protein may or may not encode polypeptide that is a BSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a BSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding a BSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 66 through 110. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a

gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 65.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a BSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 66 through 110, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H. In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a

covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS,

APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-BSP antibodies.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999), incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a BSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 66 through 110. In a preferred embodiment, the analog is one that comprises one or more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is

structurally similar to a BSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$-- and –CH_2SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of a BSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g.*, Kole *et al.*, *Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The FMOC and *t*BOC derivatives of dabcyl-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyl chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyl quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using

EDANS-FMOC-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)-β-alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-

pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is a BSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 66 through 110, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 65, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 65.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g.*, Ausubel, Chapter 16, (1992), *supra*. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences. For example, a His^6 tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. *See* Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*,

Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbrizio *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in Drosophila with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β-galactosidase, biotin trpE, protein A, β-lactamase, α-amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast α mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the BSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize BSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly BSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of BSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of BSPs.

One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65

(1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.,* Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is a BSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 66 through 110, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also

due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a BSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of a BSP may be indicative of cancer. Differential degradation of the C or N-terminus of a BSP may also be a marker or target for anticancer therapy. For example, a BSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-BSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human breast.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively

immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate

chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. *See* Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular advantage in detection of the proteins of the present invention, in human serum or tissues (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding

antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.,* Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, J. Biotechnol. 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.,* Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock et al., *J. Immunol Methods.* 231: 147-57 (1999); Young et al., *Res. Immunol.* 149: 609-10 (1998); Limonta et al., *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem.* (Tokyo) 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entireties.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, $F(ab)'_2$, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*,

United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA*.81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., Proc. Natl. Acad. Sci. (USA) 90: 7889-7893 (1993); Duan et al., Proc. Natl. Acad. Sci. (USA) 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein

fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β-galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopryanoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g.*, Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P, ^{32}P, ^{35}S, ^{3}H, and ^{125}I.

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th, ^{227}Ac, ^{225}Ac, ^{223}Ra, ^{213}Bi, ^{212}Pb, ^{212}Bi, ^{211}At, ^{203}Pb, ^{194}Os, ^{188}Re, ^{186}Re, ^{153}Sm, ^{149}Tb, ^{131}I, ^{125}I, ^{111}In, ^{105}Rh, ^{99m}Tc, ^{97}Ru, ^{90}Y, ^{90}Sr, ^{88}Y, ^{72}Se, ^{67}Cu, or ^{47}Sc.

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a BSP. In a preferred embodiment, the BSP comprises an amino acid sequence selected from SEQ ID NO: 66 through 110, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a BSNA of the invention, preferably a BSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 65, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human BSG.

The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.*, Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.*, Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.*, Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.*, Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.*, Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.*, Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.*, Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.*, Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e.*, a nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g.*, the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation

will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal

sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g.,* Gu *et al.*, *Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g.*, Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al.*, *Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g.,* Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures,

including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 65 and SEQ ID NO: 66 through 110 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting

certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis

include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Breast Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a BSNA or a BSP in a human patient that has or may have breast cancer, or who is at risk of developing breast cancer, with the expression of a BSNA or a BSP in a normal human control. For purposes of the present invention, "expression of a BSNA" or "BSNA expression" means the quantity of BSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a BSP" or "BSP expression" means the amount of BSP

that can be measured by any method known in the art or the level of translation of a BSG BSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing breast cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of BSNA or BSP in cells, tissues, organs or bodily fluids compared with levels of BSNA or BSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a BSNA or BSP in the patient versus the normal human control is associated with the presence of breast cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in the structure of the mRNA of a BSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in a BSP compared to a BSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the BSP or subcellular BSP localization.

In a preferred embodiment, the expression of a BSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 66 through 110, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the BSNA expression that is measured is the level of expression of a BSNA mRNA selected from SEQ ID NO: 1 through 65, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. BSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. BSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a BSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary,

BSNA expression may be compared to a known control, such as normal breast nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a BSP is measured by determining the level of a BSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 66 through 110, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of BSNA or BSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of breast cancer. The expression level of a BSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the BSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See*, e.g, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the BSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a BSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-BSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the BSP will bind to the anti-BSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-BSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the BSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or

more substrates are added to produce a colored reaction product that is based upon the amount of a BSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure BSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-BSP antibody is attached to a solid support and an allocated amount of a labeled BSP and a sample of interest are incubated with the solid support. The amount of labeled BSP detected which is attached to the solid support can be correlated to the quantity of a BSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a BSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more BSNAs of interest. In this approach, all or a portion of one or more BSNAs is fixed to a substrate. A sample of interest, which may comprise RNA,

e.g., total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of BSNA or BSP includes, without limitation, breast tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, breast cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary breast cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediatinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in BSNAs or BSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a BSNA or BSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other BSNA or BSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one

other cancer marker in addition to a particular BSNA or BSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a sample from a patient suspected of having breast cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP and then ascertaining whether the patient has breast cancer from the expression level of the BSNA or BSP. In general, if high expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether breast cancer has metastasized in a patient. One may identify whether the breast cancer has metastasized by measuring the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a variety of tissues. The presence of a BSNA or BSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a BSNA or BSP is associated with breast cancer. Similarly, the presence of a BSNA or BSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a BSNA or BSP is associated with breast cancer. Further, the presence of a structurally altered BSNA or BSP that is associated with breast cancer is also indicative of metastasis.

In general, if high expression relative to a control of a BSNA or BSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The BSNA or BSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with breast cancers or other breast related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of breast disorders.

Staging

The invention also provides a method of staging breast cancer in a human patient. The method comprises identifying a human patient having breast cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more BSNAs or BSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more BSNAs or BSPs is determined for each stage to obtain a standard expression level for each BSNA and BSP. Then, the BSNA or BSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The BSNA or BSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the BSNAs and BSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a BSNA or BSP to determine the stage of a breast cancer.

Monitoring

Further provided is a method of monitoring breast cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the breast cancer. The method comprises identifying a human patient that one wants to monitor for breast cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more BSNAs or BSPs, and comparing the BSNA or BSP levels over time to those BSNA or BSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a BSNA or BSP that are associated with breast cancer.

If increased expression of a BSNA or BSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a BSNA or BSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of BSNAs or BSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of breast cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a BSNA and/or BSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more BSNAs and/or BSPs are detected. The presence of higher (or lower) BSNA or BSP

levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly breast cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more BSNAs and/or BSPs of the invention can also be monitored by analyzing levels of expression of the BSNAs and/or BSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in a BSG, thereby determining if a human with the genetic lesion is susceptible to developing breast cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing breast cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the BSGs of this invention, a chromosomal rearrangement of BSG, an aberrant modification of BSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a BSG. Methods to detect such lesions in the BSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous Breast Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a sample from a patient suspected of having or known to have a noncancerous breast disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP, comparing the expression level or structural alteration of the BSNA or BSP to a normal breast control, and then ascertaining whether the patient has a noncancerous breast disease. In general, if high expression relative to a control of a BSNA or BSP is indicative of a particular noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably

the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of a noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a BSNA and/or BSP is associated with a particular noncancerous breast disease by obtaining breast tissue from a patient having a noncancerous breast disease of interest and determining which BSNAs and/or BSPs are expressed in the tissue at either a higher or a lower level than in normal breast tissue. In another embodiment, one may determine whether a BSNA or BSP exhibits structural alterations in a particular noncancerous breast disease state by obtaining breast tissue from a patient having a noncancerous breast disease of interest and determining the structural alterations in one or more BSNAs and/or BSPs relative to normal breast tissue.

Methods for Identifying Breast Tissue

In another aspect, the invention provides methods for identifying breast tissue. These methods are particularly useful in, *e.g.*, forensic science, breast cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is breast tissue or has breast tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising breast tissue or having breast tissue-like characteristics, determining whether the sample expresses one or more BSNAs and/or BSPs, and, if the sample expresses one or more BSNAs and/or BSPs, concluding that the sample comprises breast tissue. In a preferred embodiment, the BSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 66 through 110, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 65, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a BSNA can be accomplished by any method known in the art.

Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a BSP is expressed. Determining whether a sample expresses a BSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the BSP has an amino acid sequence selected from SEQ ID NO: 66 through 110, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two BSNAs and/or BSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five BSNAs and/or BSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is breast tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into breast tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new breast tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Breast Tissue

In another aspect, the invention provides methods for producing engineered breast tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a BSNA or a BSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of breast tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal breast tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered breast tissue or cells comprises one of these cell types. In another embodiment, the engineered breast tissue or cells comprises more than one breast cell type. Further, the culture conditions of the cells or tissue may require manipulation in

order to achieve full differentiation and development of the breast cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more BSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode BSPs having amino acid sequences selected from SEQ ID NO: 66 through 110, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 65, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a BSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial breast tissue may be used to treat patients who have lost some or all of their breast function.

Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a BSNA or part thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 65, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a BSP or fragment thereof. In a more preferred embodiment, the BSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 66 through 110, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-BSP antibody, preferably an antibody that specifically binds to a BSP having an amino acid that is selected from the group consisting of SEQ ID NO: 66 through 110, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example BSP polypeptide, fusion protein, or fragments thereof, antibodies specific for BSP, agonists, antagonists or inhibitors of BSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention , *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of breast function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV) , for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further

described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a BSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a BSP are administered, for example, to complement a deficiency in the native BSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid molecule encodes a BSP having the amino acid sequence of SEQ ID NO: 66 through 110, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a BSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in BSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a BSP having the amino acid sequence of SEQ ID NO: 66 through 110, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of a BSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a BSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a BSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to BSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the BSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a BSP, preferably a BSP comprising an amino acid sequence of SEQ ID NO: 66 through 110, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 65, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a BSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant BSP defect.

Protein compositions are administered, for example, to complement a deficiency in native BSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to BSP. The immune response can be used to modulate activity of BSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate BSP.

In a preferred embodiment, the polypeptide is a BSP comprising an amino acid sequence of SEQ ID NO: 66 through 110, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 65, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of BSP, or to target therapeutic agents to sites of BSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a BSP comprising an amino acid sequence of SEQ ID NO: 66 through 110, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a BSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 65, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a BSP or have a modulatory effect on the expression or activity of a BSP. Modulators which decrease the expression or activity of BSP (antagonists) are believed to be useful in treating breast cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a BSP can also be designed, synthesized and tested for use in the imaging and treatment of breast cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the BSPs identified herein. Molecules identified in the library as being capable of binding to a BSP are key candidates for further evaluation for use in the treatment of breast cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a BSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of BSP is

administered. Antagonists of BSP can be produced using methods generally known in the art. In particular, purified BSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a BSP.

In other embodiments a pharmaceutical composition comprising an agonist of a BSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP comprising an amino acid sequence of SEQ ID NO: 66 through 110, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 65, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Breast Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the breast or to specific cells in the breast. In a preferred embodiment, an anti-BSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if breast tissue needs to be selectively destroyed. This would be useful for targeting and killing breast cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting breast cell function.

In another embodiment, an anti-BSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring breast function, identifying breast cancer tumors, and identifying noncancerous breast diseases.

EXAMPLES

Example 1: Gene Expression analysis

BSGs were identified by mRNA subtraction analysis using standard methods. The sequences were extended using GeneBank sequences, Incyte's proprietary database.

From the nucleotide sequences, predicted amino acid sequences were prepared. DEX0308_1, DEX0308_2 correspond to SEQ ID NO.1, 2 etc. DEX0166 was the parent sequence found in the mRNA subtractions.

DEX0308_1	DEX0166_1	DEX0308_66
DEX0308_2	DEX0166_2	DEX0308_67
DEX0308_3	flex DEX0166_2	
DEX0308_4	DEX0166_3	DEX0308_68
DEX0308_5	flex DEX0166_3	
DEX0308_6	DEX0166_4	DEX0308_69
DEX0308_7	DEX0166_5	DEX0308_70
DEX0308_8	DEX0166_6	DEX0308_71
DEX0308_9	flex DEX0166_6	
DEX0308_10	DEX0166_7	DEX0308_72
DEX0308_11	DEX0166_8	DEX0308_73
DEX0308_12	flex DEX0166_8	
DEX0308_13	DEX0166_9	DEX0308_74
DEX0308_14	DEX0166_10	DEX0308_75
DEX0308_15	DEX0166_11	DEX0308_76
DEX0308_16	flex DEX0166_11	DEX0308_77
DEX0308_17	DEX0166_12	DEX0308_78
DEX0308_18	flex DEX0166_12	
DEX0308_19	DEX0166_13	DEX0308_79
DEX0308_20	flex DEX0166_13	DEX0308_80
DEX0308_21	DEX0166_14	DEX0308_81
DEX0308_22	flex DEX0166_14	
DEX0308_23	DEX0166_15	DEX0308_82
DEX0308_24	flex DEX0166_15	
DEX0308_25	DEX0166_16	
DEX0308_26	flex DEX0166_16	
DEX0308_27	DEX0166_17	DEX0308_83
DEX0308_28	flex DEX0166_17	
DEX0308_29	DEX0166_18	DEX0308_84
DEX0308_30	flex DEX0166_18	DEX0308_85
DEX0308_31	DEX0166_19	DEX0308_86
DEX0308_32	flex DEX0166_19	
DEX0308_33	DEX0166_20	DEX0308_87
DEX0308_34	flex DEX0166_20	DEX0308_88
DEX0308_35	DEX0166_21	DEX0308_89
DEX0308_36	flex DEX0166_21	
DEX0308_37	DEX0166_22	DEX0308_90
DEX0308_38	DEX0166_23	DEX0308_91
DEX0308_39	DEX0166_24	DEX0308_92
DEX0308_40	flex DEX0166_24	
DEX0308_41	DEX0166_25	DEX0308_93
DEX0308_42	flex DEX0166_25	
DEX0308_43	DEX0166_26	DEX0308_94
DEX0308_44	DEX0166_27	DEX0308_95
DEX0308_45	DEX0166_28	DEX0308_96
DEX0308_46	flex DEX0166_28	DEX0308_97
DEX0308_47	DEX0166_29	DEX0308_98
DEX0308_48	flex DEX0166_29	DEX0308_99
DEX0308_49	DEX0166_30	DEX0308_100
DEX0308_50	flex DEX0166_30	
DEX0308_51	DEX0166_31	DEX0308_101
DEX0308_52	flex DEX0166_31	
DEX0308_53	DEX0166_32	DEX0308_102

```
DEX0308_54  flex DEX0166_32
DEX0308_55  DEX0166_33  DEX0308_103
DEX0308_56  flex DEX0166_33   DEX0308_104
DEX0308_57  DEX0166_34  DEX0308_105
DEX0308_58  flex DEX0166_34
DEX0308_59  DEX0166_35  DEX0308_106
DEX0308_60  DEX0166_36  DEX0308_107
DEX0308_61  flex DEX0166_36
DEX0308_62  DEX0166_37  DEX0308_108
DEX0308_63  flex DEX0166_37
DEX0308_64  DEX0166_38  DEX0308_109
DEX0308_65  flex DEX0166_38   DEX0308_110
```

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'- 3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the BSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to a normal tissue (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the BSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to a normal tissue (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, BSNAs show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).
Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 65 being diagnostic markers for cancer.

Example 2B: Custom Microarray Experiment

Custom oligonucleotide microarrays were provided by Agilent Technologies, Inc. (Palo Alto, CA). The microarrays were fabricated by Agilent using their technology for the *in-situ* synthesis of 60mer oligonucleotides (Hughes, et al. 2001, Nature Biotechnology 19:342-347). The 60mer microarray probes were designed by Agilent, from gene sequences provided by diaDexus, using Agilent proprietary algorithms. Whenever possible two differents 60mers were designed for each gene of interest.

All microarray experiments were two-color experiments and were performed using Agilent-recommended protocols and reagents. Briefly, each microarray was hybridized with cRNAs synthesized from polyA+ RNA, isolated from cancer and normal tissues, labeled with fluorescent dyes Cyanine3 and Cyanine5 (NEN Life Science

Products, Inc., Boston, MA) using a linear amplification method (Agilent). In each experiment, the experimental sample was polyA+ RNA isolated from cancer tissue from a single individual and the reference sample was a pool of polyA+ RNA isolated from normal tissues of the same organ as the cancerous tissue (*i.e.* normal breast tissue in experiments with breast cancer samples). Hybridizations were carried out at 60°C, overnight using Agilent *in-situ* hybridization buffer. Following washing, arrays were scanned with a GenePix 4000B Microarray Scanner (Axon Instruments, Inc., Union City, CA). The resulting images were analyzed with GenePix Pro 3.0 Microarray Acquisition and Analysis Software (Axon). A total of 36 experiments comparing the expression patterns of breast cancer derived polyA+ RNA (9 stage 1 cancers, 23 stage 2 cancers, 4 stage 3 cancers) to polyA+ RNA isolated from a pool of 10 normal breast tissues were analyzed.

Data normalization and expression profiling were done with Expressionist software from GeneData Inc. (Daly City, CA/Basel, Switzerland). Gene expression analysis was performed using only experiments that meet certain quality criteria. The quality criteria that experiments must meet are a combination of evaluations performed by the Expressionist software and evaluations performed manually using raw and normalized data. To evaluate raw data quality, detection limits (the mean signal for a replicated negative control $\pm$ 2 Standard Deviations (SD)) for each channel were calculated. The detection limit is a measure of non-specific hybridization. Arrays with poor detection limits were not analyzed and the experiments were repeated. To evaluate normalized data quality, positive control elements included in the array were utilized. These array features should have a mean ratio of 1 (no differential expression). If these features have a mean ratio of greater than 1.5-fold up or down, the experiments were not analyzed further and were repeated. In addition to traditional scatter plots demonstrating the distribution of signal in each experiment, the Expressionist software also has minimum thresholding criteria that employs user defined parameters to identify quality data. Only those features that meet the threshhold criteria were included in the filtering and analyses carried out by Expressionist. The thresholding settings employed require a minimum area percentage of 60% [(% pixels > background $\pm$ 2SD)-(% pixels saturated)], and a minimum signal to noise ratio of 2.0 in both channels. By these criteria, very low expressors and saturated features were not included in analysis.

Relative expression data was collected from Expressionist based on meeting the quality parameters described above. Sensitivity data was calculated using an analysis tool. Up- and down- regulated genes were identified using criteria for percentage of valid values obtained, and the percentage of experiments in which the gene is up- or down-regulated. These criteria were set independently for each data set, depending on the size and the nature of the data set. Results for DEX0308_1/DEX0166_1 (SEQ ID NO:1) are shown in the following table. The first three columns of the table contain information about the sequence itself (Oligo ID, Parent ID, and SEQ ID NO), the next 3 columns show the results obtained. '%valid' indicates the percentage of 36 unique experiments total in which a valid expression value was obtained, '%up' indicates the percentage of 20 experiments in which up-regulation of at least 2.5-fold was observed, and '%down' indicates the percentage of the 36 experiments in which down-regulation of at least 2.5-fold was observed. The last column in Table 1 describes the location of the microarray probe (oligo) relative to the sequence.

OligoID	Parent ID	SEQ ID #	Sensitivity of up and down regulation			Oligo Seq location
			% valid	% up	% down	
24441	5303	DEX0308_1(SQ1)	100	38.9	0	170-222

Example 3: Protein Expression

The BSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the BSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the BSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH_2-terminus of the coding sequence of BSNA, and six histidines, flanking the COOH-terminus of the coding sequence of BSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of BSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five colum

volumes of wash buffer. BSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5'and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

The predicted antigenicity for the amino acid sequences is as follows:

DEX ID	ANTIGENICITY Position, AI Average, Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position,Max Score,Mean Score
DEX0308_100			Pkc_Phospho_Site 38-40;	
DEX0308_101			Pkc_Phospho_Site 8-10;	
DEX0308_102			Pkc_Phospho_Site 26-28;	
DEX0308_103			Myristyl 25-30; Pkc_Phospho_Site 9-11;	
DEX0308_104	12-65,1.1,54 955-969,1.09,15 931-950,1.01,20		Amidation 568-571; Asn_Glycosylation 40-43;169-172; Ck2_Phospho_Site 42-45;54-57;58-61;110-113;294-297;303-306;342-	

DEX ID	ANTIGENICITY Position, AI Average, Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position,Max Score,Mean Score
			345;408-411;426-429;447-450;619-622;713-716;734-737;776-779;830-833;941-944; Leucine_Zipper 789-810;796-817;803-824; Myristyl 27-32;82-87;212-217;318-323;469-474;968-973; Pkc_Phospho_Site 58-60;114-116;164-166;216-218;249-251;314-316;341-343;373-375;408-410;619-621;785-787;840-842;941-943; Tyr_Phospho_Site 145-151;848-855;	
DEX0308_105		1,o10-28i	Myristyl 48-53;64-69; Pkc_Phospho_Site 37-39;	
DEX0308_106			Ck2_Phospho_Site 48-51; Myristyl 70-75;	
DEX0308_107	102-114,1.22,13	2,o4-26i56-73o	Ck2_Phospho_Site 106-	27,.994,.862

DEX ID	ANTIGENICITY Position, AI Average, Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position,Max Score,Mean Score
	39-55,1,17		109; Pkc_Phospho_Site 26-28;	
DEX0308_108			Pkc_Phospho_Site 26-28;	
DEX0308_109		1,i21-43o		
DEX0308_110	264-273,1.18,109-18,1.12,10375-388,1.05,14531-565,1.03,35463-503,1.01,41446-458,1.01,13			Asn_Glycosylation 115-118;545-548;549-552; Camp_Phospho_Site 428-431; Ck2_Phospho_Site 107-110;152-155;431-434;463-466;478-481;535-538;536-539;541-544;547-550;552-555;565-568;583-586;605-608;607-610;637-640;740-743;825-828;827-830;864-867;872-875; Myristyl 4-9;103-108;137-142;179-184;302-307;456-461;498-503;528-533;531-536;576-581;601-606;754-759;814-819; Pkc_Phospho_Site 83-85;193-195;257-259;354-356;433-435;469-

DEX ID	ANTIGENICITY Position, AI Average, Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position,Max Score,Mean Score
				471;478-480;582-584;637-639;672-674;677-679;737-739;827-829; Tyr_Phospho_Site 51-57;66-74;
DEX0308_66		1,o15-34i	Ck2_Phospho_Site 46-49;52-55;71-74; Myristyl 2-7; Pkc_Phospho_Site 52-54;61-63;77-79;	
DEX0308_67			Asn_Glycosylation 42-45; Pkc_Phospho_Site 13-15; Tyr_Phospho_Site 30-36;	
DEX0308_68			Camp_Phospho_Site 28-31; Ck2_Phospho_Site 41-44; Pkc_Phospho_Site 21-23;	
DEX0308_69		1,i21-43o	Asn_Glycosylation 5-8; Ck2_Phospho_Site 47-50;	
DEX0308_70	8-25,1.05,18		Ck2_Phospho_Site 33-36; Myristyl 47-52; Pkc_Phospho_Site 15-17;18-20;48-50;	
DEX0308_71			Camp_Phospho_Site 49-	

DEX ID	ANTIGENICITY Position, AI Average, Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position,Max Score,Mean Score
			52; Ck2_Phospho _Site 55-58; Pkc_Phospho _Site 17-19;33-35;52-54;	
DEX0308_73			Ck2_Phospho _Site 9-12; Myristyl 26-31;68-73; Pkc_Phospho _Site 9-11;16-18;	
DEX0308_74	95-113,1.02,19		Ck2_Phospho _Site 16-19;99-102; Myristyl 45-50;103-108; Pkc_Phospho _Site 42-44;46-48;49-51;71-73;	
DEX0308_75	52-67,1.06,16		Ck2_Phospho _Site 8-11;58-61;	
DEX0308_76			Asn_Glycosy lation 48-51; Myristyl 37-42;42-47; Pkc_Phospho _Site 74-76;	
DEX0308_77	297-315,1.24,19 206-226,1.2,21 354-372,1.13,19 483-493,1.13,11 228-285,1.04,58		Amidation 52-55;358-361; Asn_Glycosy lation 28-31; Camp_Phosph o_Site 468-471; Ck2_Phospho _Site 4-7;30-33;58-61;64-67;81-84;98-	20,.935,.774

DEX ID	ANTIGENICITY Position, AI Average, Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position,Max Score,Mean Score
			101;136-139;273-276;279-282;398-401; Myristyl 117-122;121-126;180-185;210-215;234-239;305-310;316-321;344-349;452-457; Pkc_Phospho_Site 4-6;176-178;207-209;245-247;278-280;367-369; Prokar_Lipoprotein 225-235; Scp_Ag5_Prl_Sc7_2 201-212; Tyr_Phospho_Site 242-249;	
DEX0308_78			Amidation 42-45; Ck2_Phospho_Site 10-13; Myristyl 16-21;18-23;23-28;	
DEX0308_79	6-15,1.06,10		Pkc_Phospho_Site 42-44; Tyr_Phospho_Site 28-34;	
DEX0308_80	177-188,1.06,12 88-107,1.03,20		Atp_Gtp_A 40-47; Ck2_Phospho_Site 7-10;127-130; Myristyl 17-22;	

DEX ID	ANTIGENICITY Position, AI Average, Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position,Max Score,Mean Score
			Pkc_Phospho _Site 50-52;178-180;201-203;	
DEX0308_81			Asn_Glycosylation 8-11; Myristyl 21-26; Pkc_Phospho _Site 12-14;	
DEX0308_82	2-12,1.05,11		Myristyl 26-31;47-52;51-56;	
DEX0308_83			Ck2_Phospho _Site 52-55;	
DEX0308_84			Ck2_Phospho _Site 7-10; Pkc_Phospho _Site 13-15;	
DEX0308_85	158-189,1.12,32 259-272,1.06,14 61-100,1,40		Amidation 44-47;93-96; Asn_Glycosylation 172-175; Camp_Phospho_Site 108-111;158-161; Ck2_Phospho _Site 33-36;260-263; Glycosaminoglycan 78-81; Myristyl 10-15;73-78;100-105;112-117;177-182;227-232; Pkc_Phospho _Site 126-128;164-166;245-247;260-262;	
DEX0308_			Camp_Phosph	

DEX ID	ANTIGENICITY Position, AI Average, Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position,Max Score,Mean Score
86			o_Site 26-29;	
DEX0308_87			Pkc_Phospho_Site 5-7;12-14;	
DEX0308_88	38-50,1.12,13		Camp_Phospho_Site 18-21; Ck2_Phospho_Site 88-91; Myristyl 31-36;56-61; Pkc_Phospho_Site 24-26;99-101;106-108;	
DEX0308_90	47-57,1.23,11		Pkc_Phospho_Site 20-22;48-50;	
DEX0308_91			Ck2_Phospho_Site 24-27;	
DEX0308_92		1,i7-29o	Asn_Glycosylation 42-45; Pkc_Phospho_Site 31-33;	
DEX0308_93			Amidation 33-36; Camp_Phospho_Site 19-22; Ck2_Phospho_Site 4-7;40-43; Pkc_Phospho_Site 33-35; Tyr_Phospho_Site 35-42;36-42;	
DEX0308_94	35-57,1.17,23		Ck2_Phospho_Site 42-45; Myristyl 5-10;9-14;64-69;68-73;124-129; Pkc_Phospho_Site 36-	

DEX ID	ANTIGENICITY Position, AI Average, Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position,Max Score,Mean Score
			38;42-44;95-97;101-103;	
DEX0308_95			Myristyl 2-7; Pkc_Phospho_Site 20-22;	
DEX0308_96	21-33,1.15,13		Pkc_Phospho_Site 51-53;67-69;	
DEX0308_97	221-243,1,23		Amidation 195-198; Camp_Phospho_Site 197-200; Ck2_Phospho_Site 24-27;69-72;89-92;178-181; Myristyl 144-149;148-153; Pkc_Phospho_Site 89-91;94-96;192-194;214-216;228-230;281-283; Tyr_Phospho_Site 7-14;197-205;198-205;	
DEX0308_98	17-26,1.02,10		Myristyl 26-31; Pkc_Phospho_Site 2-4;10-12;31-33;	
DEX0308_99	108-136,1.06,29		Amidation 64-67;74-77;109-112; Ck2_Phospho_Site 133-136;	

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 65. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al., Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al., Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al., Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Manheim), and FISH is performed as described in Johnson *et al., Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and

translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 μg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 μl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 , μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and

EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about

3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 μg/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5'and 3'end sequences respectively as set forth in Example 1. Preferably, the 5'primer contains an EcoRI site and the 3'primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the

presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et

al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They

may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 μg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989); etc. For a review of such techniques, see Gordon,"Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated

immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications

to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the

cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 66 through 110;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 65;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule having at least 60% sequence identity to the nucleic acid molecule of (a) or (b).

2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.

3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.

4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.

5. The nucleic acid molecule according to claim 4, wherein the nucleic acid molecule is a human nucleic acid molecule.

6. A method for determining the presence of a breast specific nucleic acid (BSNA) in a sample, comprising the steps of:

(a) contacting the sample with the nucleic acid molecule according to claim 1 under conditions in which the nucleic acid molecule will selectively hybridize to a breast specific nucleic acid; and

(b) detecting hybridization of the nucleic acid molecule to a BSNA in the sample, wherein the detection of the hybridization indicates the presence of a BSNA in the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 66 through 110; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 65.

12. An antibody or fragment thereof that specifically binds to the polypeptide according to claim 11.

13. A method for determining the presence of a breast specific protein in a sample, comprising the steps of:

(a) contacting the sample with the antibody according to claim 12 under conditions in which the antibody will selectively bind to the breast specific protein; and

(b) detecting binding of the antibody to a breast specific protein in the sample, wherein the detection of binding indicates the presence of a breast specific protein in the sample.

14. A method for diagnosing and monitoring the presence and metastases of breast cancer in a patient, comprising the steps of:

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the breast specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of breast cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient.

16. A method of treating a patient with breast cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the breast cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

SEQUENCE LISTING

<110> diaDexus, Inc.
Salceda, Susana
Macina, Roberto
Hu, Ping
Recipon, Herve
Karra, Kalpana
Cafferkey, Robert
Liu, Chenghua
Sun, Yongming

<120> Compositions and Methods Relating to Breast Specific Genes and Proteins

<130> DEX-0308

<150> 60/268,289
<151> 2001-02-13

<160> 110

<170> PatentIn version 3.1

<210> 1
<211> 999
<212> DNA
<213> Homo sapien

<400> 1
ggataacaac cgaaagtgat tatatatggg ccatgggtct ctagatcatg ctcgagcgcg 60

cgcagtgtga tggatgcggc gcccgggcag gtactttgtc cctgattaaa taatgtgacg 120

gatagcaatg catcaagtgt ttattatgaa aagagtggaa aagtatatag cttttagcaa 180

aaggtgttgg cccattctaa gaagatgagc gaatatatag aagatacgtg tgggcatttc 240

ttcctgttag gtggagctgt atgctgttga cgtttctccc catactcttc ccactctgtt 300

ttctccccat tatttgaata aagtgactgc tgaagatgac ttggaatcct tatccactta 360

gatttaatgt ttagagaaaa acctgtaggt ggaaagtaag actccttccc tgaattgtca 420

gtttagagca acttgagaga agagtagaca aaaaataaaa tgcacataga aaaagagaaa 480

aagggcacaa agggattggc ccaatattga ttctttttttt ataaaacctg cctttggctt 540

agaaggaatg actctagcta caataataca cagtatcgtt caagcaggtt cccttggttg 600

ttgcattaaa tgtaatccac ctttaggtat cttagaacca cagaacaaac actgtgtttg 660

atctagtagg tttctatttt tccttttctct ttacaatgca cataatactt tcctgtattt 720

atatcataac gtgtatagtg taaaatgtga atgacttttt tcgtgaatga aaatctaaaa 780

tctttgtaac tttttatatc tgcttttgtt tcaccaaaga aacctaaaat ccttctttta 840

aaaaaaaaaa caaaaaaaca aaaaaaaaaa aaggcggggg gtacccaggg gccaaagctg 900

gcgccggggg ggacattggt ttcccggccc acattccccc ccatatcgca caaaaaaaag 960

ggacaggaga gcgagccaag aagaaccaac cagagaaag 999

<210> 2
<211> 557
<212> DNA
<213> Homo sapien

<400> 2
actctaatat aaaggacagg tggtgtttct aaataattgg ctgctatggt tctgtaaaaa 60

ccagttaatt ctatttttca aggtttttgg caaagcacat caatgttaga ctagttgaag 120

tggaattgta taattcaatt cgataattga tctcatgggc tttccctggg aggaaaggtt 180

ttttttgtgg tgtttttttt aagaacttga aacttgtaaa ctgaagatgt ctgtgagctt 240

ttttgcccat ctgtaggtgt actgtgaaga tttcaaaacc tgagagcact tttcttgtg 300

tgttagaatt atgagaaagt ggctagatga ctttaggatt tgcgattttt ccctttattg 360

gctcatttct ttgtgacgcc tttgtttggg gagggaaatc tgtttatttt ttcctacaaa 420

taaaaagcta agattctata tcgcaaaaaa aaaaaaaaaa aaaaaaaaaa aaggtggggg 480

gaaactcggg gcaaaagggg tccccggggg gaaattggtt ttcggtcaaa attcccaaat 540

attagaaaaa aaaaaga 557

<210> 3
<211> 1200
<212> DNA
<213> Homo sapien

<400> 3
atggcgtggc ggcggcgcga agccggcgtc ggggctcgcg gcgtgttggc tctggcgttg 60

ctcgccctgg ccctgtgcgt gcccggggcc cggggccggg ctctcgagtg gttctcggcc 120

gtggtaaaca tcgagtacgt ggacccgcag accaacctga cggtgtggag cgtctcggag 180

agtggccgct tcggcgacag ctcgcccaag gagggcgcgc atggcctggt gggcgtcccg 240

tgggcgcccg gcggagacct cgagggctgc gcgcccgaca cgcgcttctt cgtgcccgag 300

cccggcggcc gaggggccgc gccctgggtc gccctggtgg ctcgtggggg ctgcaccttc 360

aaggacaagg tgctggtggc ggcgcggagg aacgcctcgg ccgtcgtcct ctacaatgag 420

gagcgctacg ggaacatcac cttgcccatg tctcacgcgg gaacaggaaa tatagtggtc 480

attatgatta gctatccaaa aggaagagaa attttggagc tggtgcaaaa aggaattcca 540

gtaacgatga ccataggggt tggcacccgg catgtacagg agttcatcag cggtcagtct 600

gtggtgtttg tggccattgc cttcatcacc atgatgatta tctcgttagc ctggctaata 660

ttttactata tacagcgttt cctatatact ggctctcaga ttggaagtca gagccataga 720

aaagaaacta agaaagttat tggccagctt ctacttcata ctgtaaagca tggagaaaag 780

ggaattgatg ttgatgctga aaattgtgca gtgtgtattg aaaatttcaa agtaaaggat 840

attattagaa ttctgccatg caagcatatt tttcatagaa tatgcattga cccatggctt 900

ttggatcacc gaacatgtcc aatgtgtaaa cttgatgtca tcaaagccct aggatattgg 960

ggagagcctg gggatgtaca ggagatgcct gctccagaat ctcctcctgg aagggatcca 1020

gctgcaaatt tgagtctagc tttaccagat gatgacggaa gtgatgagag cagtccacca 1080

tcagcctccc ctgctgaatc tgagccacag tgtgatccca gctttaaagg agatgcagga 1140

gaaaatacgg cattgctaga agccggcagg agtgactctc ggcatggagg acccatctcc 1200

<210> 4
<211> 816
<212> DNA
<213> Homo sapien

<400> 4
accactctac cctccgcacc tcctcctgca tcagccggcc tgaagtcgca ccctcctcct 60

ccggatgaag tagagaaata aatttctccc accctaaacc agtctttgag ctgattgcag 120

tatgactcca tttaccctgc tgcattcata taatagttca cctggtgcaa aacaactgaa 180

gattatttac aatgctaccc tgctttttct ggtgtcctga acctgcgaag ttgtgctttt 240

taacgtctta tgatgtaatc agcgcgattt cacttacctg aatttcgcat gaattctaca 300

gacatgggca agatcgggtt gtaagacctc tgagatttaa ggccatgccc ctggatcatg 360

gtgaacttac caaagcaaac aatgcctgtg agatggtcct gcagcagcca accagtgaac 420

tcttttggtg acatccgtgt tcttgttgta taactttata ttcctataaa tccattaagg 480

ccccaataaa gtttgtctct aagcgctgtg ttagatctat atgactacat ctagtaaatt 540

gtgaatttta agtaaatatt ttataagaac tcctatgtaa agcattacta aaattagtgt 600

tgaaatatga ccttcttcct acatttattc atttatttat gtctatttat tcatttattt 660

tagtgaaaaa tataaggcaa agtagaggaa ggttcaaatc cgaaaaaaaa aaaaaaaaaa 720

aaaaaaaaag cgctgggggt acctctgggc caaaggggtc ccgggggggaa ttggtttccc 780

gccccaaatt ccccccccaac tttccgccca agggtc 816

<210> 5
<211> 1029
<212> DNA

<213> Homo sapien

<400> 5
accactctac cctccgcacc tcctcctgca tcagccggcc tgaagtcgca ccctcctcct 60

ccggatgaag tagagaaata aatttctccc accctaaacc agtctttgag ctgattgcag 120

tatgactcca tttaccctgc tgcattcata taatagttca cctggtgcaa aacaactgaa 180

gattatttac aatgctaccc tgcttttttct ggtgtcctga acctggaagt tgtgcttttt 240

aagtcttatg atgtaatcag cgcgatttca cttcctgaat ttcgatgaat tctaagacat 300

gggcaagatc gggttgtaag acctctgaga tttaaggcca tgccctggat catggtgaac 360

ttaccaaagc aaacaatgcc tgtgagatgg tcctgcagca gccaaccagt gaactctttt 420

ggtgacatcc tgttcttgtt gtataacttt atattcctat aaatccatta aggccccaat 480

aaagtttgtc tctaagcgct gtgttagatc tatatgacta catctagtaa attgtgaatt 540

ttaagtaaat attttataag aactcctatg taaagcatta ctaaaattag tgttgaaata 600

tgaccttctt cctacattta ttcatttatt tatgtctatt tattcattta ttttagtgaa 660

aaatataagg aaagtagagg aaggttaaat ccaaaaaaga attgtttcca gtacactttc 720

tttaatttgc tgtcagtttt tgcatggaat ctacatcttt ttatgctaat cctcatccta 780

gtattttaca tcttaactat ttttttctga ctgaaatggt tgatgtgctt gtttttgta 840

attttctact ttccttctaa aatgcttagt attgaacaaa tagaatatcc taattaaaaa 900

cagtaataaa tattatggtg aaaaaataca agtaaaatgg gaaaacatta gatagcagct 960

ttcaatattt catatagttc ataaatgttt caggaattac aaggttatag aaaaaaattt 1020

atagactat 1029

<210> 6
<211> 811
<212> DNA
<213> Homo sapien

<400> 6
gaagatccac atagggctgg gtcctctaga tgctgctcga gcggcgcagt gtgatggatg 60

cgtggtcgcg gcgaggtaca aataattctt ttatgaaaaa taaaactcta cttatgcata 120

cctggttgac aatatgacaa ttttaaacta cagtataaat atgagatgtt ggttaaaatc 180

cttcagcagg cttcttatgt ctactagtgt tctagtcttt cttggcacat cctatttcta 240

tttaggcttc tggccctacc tctctagcat cacttctcct gaaaccagcc atgggaactg 300

aaacaactaa agaatgtgtc aagtacacta gaacggaaat taaagctgct aacattctaa 360

gccattagac ctatattatt ctctgtgtgt gtgcacatgt gtgtatcgga tctgactatc 420

tgactgtgtg taactatgta taacgaatat tcgactcttc acccacttaa ctctgaccaa 480

aataacgctg cacttaaaaa gtatcccaaa acttactggc ttaaaacgct gacatcagtt 540

atccaacaga tcttcagatt ggctgacatt tgtccaaagt cagtcttgca tggatggttc 600

taactggtct ctctcattca tactctggaa ccagtttgag ttcacttggg cagtggctct 660

gcctcacatg ttgcatatcc tcctgtggga ccagcagact agtctaaagc atatccttct 720

tgtgctacca taaggttcaa aagtaagctt tataaacttc tgttcatgtc ccgtctgcta 780

atattccatt gcctctccca gaagactgct a 811

<210> 7
<211> 869
<212> DNA
<213> Homo sapien

<400> 7
agcgccgcca gttgtgatgg atggcagccc gggcaggtac cctaacctga gggggccacc 60

acacccaggc ccacaaactt gatctcagtg gtaactcctg tcctttctgt cccatgagcc 120

acattctgaa cagcctgatc aggatcctca accgtcaggc tcactaagat ccgagcaaca 180

ttttccttcc ttttgttagt tttatgggtt gttttggtgt ctggggtttt tacacaaaaa 240

aaaacactca tttgatattg gcatgaacag agatggctgc aatttttatt ctcttgggag 300

tgttctattg atacaatgtt ttaatttttc agcttgacca tcttgcctct ttgagaagag 360

agagaagtgg gcatccttcc tttaaattca ggaaccactg gtggttttat ttggactttt 420

tctggttact ggcatcccct atataagtgg tttgggattc ggggactatg tctcgggggg 480

agaaaaactc ccagttagtt cgtgtattgg gtatgggtta ttcagcttac tttgggtatc 540

aaaattattg ccagttttag agctcacttg agctgaagtt tatcgtcaca agattctgtt 600

taacatgctt tccttgtttg tggaaacaag caaaaacttc ccttttgtg ttacgggatt 660

tgtgacctac aaatcctaat catgtttaaa atgtgccggt gtcgggtaga tgacttttct 720

gccctctggg ggtcaccttt attatttaag gataccttta aattacaaca aacacaacaa 780

caccagatca ccaaacacac acggcgcggg gacccgggcg acaacgcggc ccccgggga 840

aaagtgtccg gcccaatcaa gtgtgagga 869

<210> 8
<211> 883
<212> DNA
<213> Homo sapien

<400> 8

actgtgggaa ggggagttgg gcactcttgg aggagctcct gctgaaggtg gtcagcctgc 60

ctgacaatgg aaggcatact tgaatgggga gcagggtatg tgctttcata tgaaaaaaga 120

gctgatgtta aaactcattt ggtgaggtca acgttgtcac ataccttcac ataagggata 180

gtatatttta gggttgcagt caaacttgtg ctcagcactg gtgaaactga gagtcaggct 240

tttacatttt aaagaaaata cagtttacat ctctaattca ggtgtctact tattttatgt 300

gggaataata tttagatttc ccccccacca tgaaggtttc ttcctatttt ttatagtgcg 360

tgtaactttc acccccaatc tttatctctg gatttttca ctctttaaat ttggaagttg 420

actagcattt tcaaaccttt attttatacc ctgtgtcttt tatattaact ttttcttatt 480

attctttagg taagaatgat tagatgttgg ctgatatagg agtgctcatt cacatgaagt 540

ggatagatac ttcctcaaga catcacacag cggtgcagtc aatccaaggc agggaagcca 600

caagcagact gacaacgttt ctagcaggat caggtgagct gtgtccaaga aaaccaacga 660

gaaggagtgg aacggaggaa tgaacgtttc attctcgtta ataaaggcat tatcctaatc 720

aaaaaaaaaa aaaaaaaaaa aaaggcttgg gggtacccag ggccaaagcg gttcccgggg 780

tgacacttgg ttacccgctc caaaatttcc acacaccttc cgcgcaccac acggaaaaca 840

aacaagacga aagaaccaga agaaacacaa aaaaataaga ata 883

<210> 9
<211> 2898
<212> DNA
<213> Homo sapien

<400> 9
ggccattatg gccgggagtg atgtcagcta gtgcagttct caaatggctg cctattaggg 60

aaagaattca gaggatttga ctgctcctaa tcatctgtca ttgctgctag ataatgattg 120

gcaattttta agactcaact ggaaatctca acagttgctg gtaaaccatt aaccataaaa 180

acgttgcttt tgaacaccag tgctgaaaaa aatatttttt ttttttttt gagagtgaaa 240

agggcttgga cttaagatag gacaatgtgg agaatggggg gaagaatgca aaacgatata 300

gtatcccttа tggatggtac atgtgcaaca gggaactctt acttcatata ccctttgcag 360

taatcattca gggaggaaga aaaacctgga acttgaatga aggctgatct ttgttttgtg 420

cactgtggcc ctgccaggca tatagtgaag gtgaatgtct tctccctcag aaaaaaattg 480

gttccttgct gtcccagtaa ggcatagctt ttccagccct aactttaaaa ctcagtgagg 540

acttagatgg gaaagaatga ggtaaataca aaggattgca ggacaacaac tacagcgttg 600

tgtactgtgg gaaggggagt tgggcactct tggaggactc ctgctgaagg tggtcagcct 660

gcctgacaat ggaagacata cttgaatggg gagcagggta tgtgctttca tatgaaaaaa 720

gagctgatgt taaaactcat ttggtgaggt caacgttgtc acataccttc acataaggga 780

tagtatattt tgggttgcag tcaaacttgt gctcagactg gtgaaactga gagtcaggct 840

tttacatttt aaagaaaata cagttttcat tctaattcag gtgtctactt attttatgta 900

agaataattt tagatttccc ccccaccatg aagtttcttc ctatttttt tatgctgtaa 960

cttaccccca atctttatct ctggattttt actctttaaa ttttgaagtt gactagcatt 1020

ttcaaacctt tattttatac ccttgtcttt tatattaact ttttcttatt attctttagg 1080

taagaatgat tgatgttggc tgatattgga gtgctcattc acatgaagtg gatagatact 1140

tctcaagaca tcacacagcg tgagtcaatc aaggagggaa gccacaagca gactgacaac 1200

gtttctagca ggatcaggtg agctgtgtcc agaaaaccaa cgagaaggag tggaaggagg 1260

aatgaacgtt tcattctcgt taataaaggc attatcctaa ttaaaaaaaa aaaaaaaaaa 1320

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaagata 1380

aaaaaaaaaa aaaaataata ataaaaaaaa aaaaaaaata aaaaaaaaaa aaaaaaaaaa 1440

aagaaaaaag aagagagcaa gtaggctata taatagttaa attggagaat gtggtatttt 1500

tggaatgata taagagaaaa tcagagagac ggagcgaaca cacaaagctg ggaacagcca 1560

gaccaacact aaagacgaaa gtaaggaaga caacgacata agggcgacaa acgtacacac 1620

aaacccccaa gccactaaga aacaaaaaag gaatgagaag aaaacacaga agactacaca 1680

acagctatgc gcccaagcag aatgcactaa accagacaca ctaacgcaac acatcaaccg 1740

aaaacaaaag agagaaatag cggacagaaa gagagagatc aatatcagaa cagcccaacg 1800

caaagagcta gatgatgcaa ccaaacctag acacgaacaa tcagtgagtg atgaaaaaca 1860

tcacaacaac acgagcactg aaaccgacat aatcaaaaac aaacgaaaca acacgactaa 1920

tacaggacgg aacacctaga cgcacgacga caacaaacac tcaacacgaa acaccagcac 1980

ccaacagatg cacagaaatg acaacaaacc agaccggaga caagaaatca taatactaga 2040

aaaagaaaaa cacataaact tatcacacaa atcacctaca cataaaacat aacgacaaat 2100

acaaaatact aaataaaaaa ataatctaca acacacaata aaaccaataa aacaacaatc 2160

acacacacat ctagaccata tacacattat acaaacacaa tatatatcta tatcaaatca 2220

agacaaaaac acatacaaat tacaaaatac aacactaaag aagactataa catcaatata 2280

atatatcaat aacgaaatca acagtaacac cagttaaaca atacatatca caagaaacac 2340

aactacgaaa gcagagaaga cgataggaga gagagaagag agagatgaac gagagcgacg 2400

agaaacagga cgagaggccg aaattagatg gcagaggcgc gaacgctgca gaagcggaag 2460

ggagagcaga gaaaatagag tgggcgggat gacagaggta ggcaacagag gaggatgagc 2520

gaggaagaag cgaagtcgag acggcacgaa acgcaggatg cagtaacgac tgacacacga 2580

ggaggcagac cagacagatg agcagcgcga gagcgaacga ccagcactca gatgcgaccc 2640

agacggagaa agcgacgaag ggcagagcga gacgagcgag cgagagcgcg atcgcaaacc 2700

tacagatcat ctcgtcgagc acaacacgac gaaggcgcga tcgagatgca tagacgcgac 2760

tgcgagcaca acggcccgga gaaccggagc gcacaagcga ggtcggatga gagcaacaga 2820

attgagcttg gaggatagag tgagaaaaag aaagaacgaa caaaccgaca tcccagcaca 2880

acaacacacc aaaaaaca 2898

<210> 10
<211> 810
<212> DNA
<213> Homo sapien

<400> 10
gcgtggtcgc ggccgaggta cttaccatgt tctgttctga gaatactctg cctcaagata 60

tcctacaact atcttactgt attcagctct ctgctcaagt attaactgat gaaacctgtc 120

atccctactc cactccatgt tctgctttac ttaacagcaa tgcacatatg gccccctgaa 180

taatatacat ttagtcactt atttttactt atctgctaat taaaatgtag acttttttcta 240

ttctgtttac tgctgtattc ccagcatgtt ttatccgaat gtgcagggtt tcttttcttc 300

tcccttatcg tgggaagtga tgtgcacaaa tacacataat ggagcctgaa tgtcatatgc 360

tttcatacct gtgtgaattc tggtaagaaa ggaaaagtag cgatgacagg taatataatt 420

acattaagtc actctcatag ttagctgttt attgctttcc tgctcttatt ctcagtcccc 480

aggaccaaat gttgaccact accttccccc acatataatt aggttattta ccgaacgcca 540

tgcaggtggc tgttaaaagg aagatatata cttaccttat aaactcaact tttccctgtt 600

gtctttctgt ctcaccccta cctccatgct ttaaattaac ttttcaggct taggccttat 660

ctctcagtag agccatatca ggtatgtgtc aaagccggaa atgtttcctg gggatgagct 720

ggggtatcat ggtcatagct gttcctgtgt gaattgtttc gtcacatcca ccccggccag 780

gaggggtaaa gaaaaaaaga aaaaaaaaaa 810

<210> 11
<211> 889
<212> DNA
<213> Homo sapien

<400> 11

tacccatcag gaatatccgg ctcagaagcc aggtcctcag agactgttct ctcactgagg 60

acctagagag ggaactccct gtgttattct cactgatggc ccaggaacca cccttgcaag 120

tcatgaccac cagcatcatg tagcactgga actgatctgg gcaatgacct ctgtctaaac 180

ttctgaatcc cctccgacaa agacccaaga cagcagcatg gccatgcagc tgtgctcaca 240

tctcaccccc gcactggcca ggaacacatc tatctttcct ttggggtagg gtcacccaac 300

tggctctggc cacttccagc gtgtgaaagg catctatgtg acagacccct ctgcagtttg 360

aaactgtgtg acaatcttta acacccaact cagcatctgc atgcggtttc tgagaattac 420

ctatatcttt tgtggtctct ttgctgattc tctgtttcat taaaaaaaaa aaaaagagtg 480

actcggtgat cccgtgagtt tcctatatag ccaattttac tcactagcta aagaaacact 540

tgtatttcaa aatgaccaaa cctagccaac aattaggcaa gctctcatca ggactccatg 600

cagggcctgt gtgattgcct aaaaaaagtc ttccacagcg gatcttgaac ttggaccatg 660

gggggctgtt gcccacattg aacctcaggg ggctgatggg cgagaacgca ggaaggagtc 720

acacacattg gaactgtaga aaatacccca tgtgggtgga attcatcacc caaagaagaa 780

attcctgtaa cctacttggt gcttgtgtgt gccgggatgg ggagtcccgg cccaggaatt 840

caagtgtctc ttcaagtggg gtacaggcaa gcggtctggt gaaggatcc 889

<210> 12
<211> 1572
<212> DNA
<213> Homo sapien

<400> 12
ttaactactt tatcgacctt caaatggcct aagcaattaa gttccaattt ctttaacctc 60

acactcaaag cctttacaat ttagttttca actgccttcc tatacttccc catcttccac 120

cctttaagtc ctgtatctac tcacagtttt ccacacctac cctgaatttc cccactttag 180

tttcattaat agttttgtca ctgcaatgac agaactgtta aagcccagct taaatttatt 240

taaaagttta caagttgttc tgggaatcat atagtttgat cctcagtagt gataaaacaa 300

cataaaatta tgaaaaatgt tattataaca taatggaatt tcctctactt taaatattta 360

ttttgcacca tccctgacct cactaccaaa aaaaaaaaaa ttcaaagtgc ctgaggtttc 420

caggcattct tagctctatt tacttacttc ccacctcaaa tggccttaga attcaaattc 480

tgtagaaaat ggattgccat aaataatcca atgaaaatgg gtcatatttt gccattaata 540

gaatcacagt caacaaggac taatagaatt agtcacttaa gtatctttag atatgggaga 600

caacagaaac aataagaatc tcttctcttt gttcccagcc ttgaatataa ctaggaagcc 660

ttcccagaag aaagcagctg tgaagggtac ccatcaggaa tatccggctc agaagccagg 720

gcctcagaga ctgttctctc actgagacct agagagggaa ctccctgtgt tattctcact 780

gatggcccag gaaccaccct tgcaagtcat gaccaccagc atcatgtagc actggaactg 840

atctgggcaa tgacctctgt ctaaacttct gaatcccctc cgacaaagac ccaagacagc 900

agcatggcca tgcagctgtg ctcacatctc acccctgcac tggccaggaa cacatctatc 960

tttcctttgg gtagggtcac ccaactgctc tgccacttcc agcgtgtgaa aggcatctat 1020

gtgacagacc cctctgcagt ttgaaactgt gtgacaatct ttaacaccca actcagcatc 1080

tgcatgcggt ttctgagaat tacctatatc ttttgtggtc tctttgctga ttctctgttt 1140

cattaaaaaa aaaaaaaaga gtgactcggt gatcccgtga gtttcctata tagccaattt 1200

tactcactag ctaaagaaac acttgtattt caaaatgacc aaacctagcc aacaattagg 1260

caagctctca tcaggactcc atgcagggcc tgtgtgattg cctaaaaaaa gtctttcaca 1320

gcggatctag aacttggacc atgggggget gttgcccaca ttgaacctca gggggctgat 1380

gggcgagaac gcaggaagga gtcacacaca ttggaactgt agaaaatacc ccatgtgggt 1440

ggaattcatc acccaaagaa gaaattcctg taacctactt ggtgcttgtg tgtgccggga 1500

tggggagtcc cggcccagga attcaagtgt ctcttcaagt ggggtacagg caagcggtct 1560

ggtgaaggat cc 1572

<210> 13
<211> 665
<212> DNA
<213> Homo sapien

<400> 13
cgtgataacc agataggcga tgcgcctcta gatcatgtcg acgcggcgcc agtgtatgga 60

tagcggacgt cggagaggta ctctggggag tgccatcatt tgtggtctct gcccagagat 120

cggagtaaca gctgatccag ctgacacgta tccagctact ggtcctgctg atgatgaagc 180

ccctgatgct gaaaccactg tctgctgcaa ccactgcaac cactgctgct cctaccactg 240

tcaaccaccg gctgcttcta ccactgcctc gtaaagacat tccagtttta cccaaatggg 300

ttggggatct cccgaaatgg tagagtgtgt ccctgagaat ggaatcagct tgagtcttct 360

gcaattggtc acaactattc atgcgttcct gtgatttcat ccaactacgt taccttgccg 420

tacgatatcc ccattgtagt ctcgtaatca gtttattttc tttcaaataa aaaataacta 480

tgagcaacaa aaaagaaaaa acaaaaaaaa aaacaaaaaa aaaagcggtc gggggtacc 540

tcaggggcca aacgctggtt cccgggggta gaactgggta cccggctcac aatcccacca 600

cacctcgcag cacagaggcg agcacgggaa acacacacga cgcgcaagga agcggccgca 660

cgcac 665

<210> 14
<211> 762
<212> DNA
<213> Homo sapien

<400> 14
ggatgagtag atcactatag ggcgactggg ttctctaatg ctgctcgagc ggcgcagtgt 60

gatggatccg cccgggcagg tacttgaaaa tgaatgaatg gcttcccgag aggcagaagg 120

caggggggtgt gccctacccc acgccggcca agagttcaac aagcattggt tgacaagtga 180

atagtgagca cttgaaccca gtcacaattc aagatgaggg ctctgccatg acgcatgtgg 240

tctgtgtcac cctgcagtct ccctgagcag tgtctgaggt tcgagtggga ccctacattc 300

gtgaacgaga tttatcatct ccccaggcaa aataacagat tctgtcctag gtgttgtgat 360

gtaacaatgg tagcgatcac agccataact tacaattatt ggcatactta cgacgagtcc 420

cgcactgggc ctaagtgctt tttaactatg tgaaatgttt ctttccttga ttgatgccaa 480

acatgaataa agataatttt ctgtatctgc taaaaaaaaa aaaaaaaaaa aaagaaaaaa 540

agggggggga cactaaggtg gaattttaaa ggggatcccc tatttttgtt tacaatcttt 600

ttttttctgg agggtaatta aatttaacga ggggtttttg aaaggtgccc tcaaaaaaaa 660

aaatgaaaaa aaaaaaaaag cgtgggggggg tgaacggggc ataaaaggtt tcccgggtgg 720

aaaattggtt tccggggcac aaattcaaga aacaaaaaaa ga 762

<210> 15
<211> 930
<212> DNA
<213> Homo sapien

<400> 15
ccgcccgggc aggtggcgcc tggttctgcg cgcaccggct gtacggagca ggagcaagag 60

gtcgccgcca gcctcagccg ccgagcctcg ttcgtgtccc cgcccctcgc tcctgcagct 120

actgctcaga aacgctgggg cgcccaccct ggcagactaa cgaagcagct cccttcccac 180

cccaactgca ggtctaattt tggacgcttt gcctgccatt tcttccaggt tgagggagcc 240

gcagaggcgg aggctcgcgt attcctgcag tcagcaccca cgtcgccccc ggacgctcgg 300

tgctcaggcc cttcgcgagc ggggctctcc gtctgcggtc ccttgtgaag gctctgggcg 360

gctgcagagg ccggccgtcc ggtttggctc acctctccca ggaaacttca cactggagag 420

ccaaaaggag tggaagagcc tgtcttggag attttcctgg ggaaatcctg aggtcattca 480

ttatgaagtg taccgcgcgg gagtggctca gagtaaccac agtgctgttc atggctagag 540

caattccagc catggtggtt cccaaatgcc actttattgg agaaactttt tggaacaata 600

catggatgag gatggtgagt ggtggatagc caaacaacga gggaaaaggg ccatcacagc 660

acaatgacat gcagagtatt ttggaccttc ataataaatt acgaagctca ggtgtatcca 720

acagcctcta atatggagta tatgacatgg gatgtagagc tggaaaggat ctagcagaga 780

tcctggggct gaaagttgcg ttgtggggaa cagtgtgacc tgcgaggcta tggggtagtc 840

ataggagcga taggtgtttc ctagtgtgaa gattggtgat cgcgcgtcga caataacgca 900

gaaaacgata gagaggagag gagaagagag 930

<210> 16
<211> 1500
<212> DNA
<213> Homo sapien

<400> 16
atgaagtgta ccgcgcggga gtggctcaga gtaaccacag tgctgttcat ggctagagca 60

attccagcca tggtggttcc caatgccact ttattggaga aacttttgga aaaatacatg 120

gatgaggatg gtgagtggtg gatagccaaa caacgaggga aaagggccat cacagacaat 180

gacatgcaga gtattttgga ccttcataat aaattacgaa gtcaggtgta tccaacagcc 240

tctaatatgg agtatatgac atgggatgta gagctggaaa gatctgcaga atcctgggct 300

gaaagttgct tgtgggaaca tggacctgca agcttgcttc catcaattgg acagaatttg 360

ggagcacact ggggaagata taggcccccg acgtttcatg tacaatcgtg gtatgatgaa 420

gtgaaagact ttagctaccc atatgaacat gaatgcaacc catattgtcc attcaggtgt 480

tctggccctg tatgtacaca ttatacacag gtcgtgtggg caactagtaa cagaatcggt 540

tgtgccatta atttgtgtca taacatgaac atctgggggc agatatggcc caaagctgtc 600

tacctggtgt gcaattactc cccaaaggga aactggtggg gccatgcccc ttacaaacat 660

gggcggccct gttctgcttg cccacctagt tttggagggg gctgtagaga aaatctgtgc 720

tacaaagaag ggtcagacag gtattatccc cctcgagaag aggaaacaaa tgaaatagaa 780

cgacagcagt cacaagtcca tgacacccat gtccggacaa gatcagatga tagtagcaga 840

aatgaagtca taagcgcaca gcaaatgtcc caaattgttt cttgtgaagt aagattaaga 900

gatcagtgca aaggaacaac ctgcaatagg tacgaatgtc ctgctggctg tttggatagt 960

aaagctaaag ttattggcag tgtacattat gaaatgcaat ccagcatctg tagagctgca 1020

attcattatg gtataataga caatgatggt ggctgggtag atatcactag acaaggaaga 1080

aagcattatt tcatcaagtc caatagaaat ggtattcaaa caattggcaa atatcagtct 1140

gctaattcct tcacagtctc taaagtaaca gttcaggctg tgacttgtga aacaactgtg 1200

gaacagctct gtccatttca taagcctgct tcacattgcc caagagtata ctgtcctcgt 1260

aactgtatgc aagcaaatcc acattatgct cgtgtaattg gaactcgagt ttattctgat 1320

ctgtccagta tctgcagagc agcagtacat gctggagtgg ttcgaaatca cggtggttat 1380

gttgatgtaa tgcctgtgga caaaagaaag acctacattg cttcttttca gaatggaatc 1440

ttctcagaaa gtttacagaa tcctccagga ggaaaggcat tcagagtgtt tgctgttgtg 1500

<210> 17
<211> 296
<212> DNA
<213> Homo sapien

<400> 17
acagagttct tatgtgtgtg agttttctat ggtgactaca caaaacctca ggcttacaat 60

tgtggaggtc agaggtcaag gtgctggcag ggcaggatcc ttcctttcct ccatcatggg 120

ggctgctggc agaattcagt ttcttgcagg gctgggacgg aggtccccag tcccagctgc 180

ttaggggcca ccacactcct cggccctcct ctaaggccag cagcgcaggt gcggccctcc 240

tcgggttcta acctctcctg cttctggcat ctctcagact cagcaggaaa ggctct 296

<210> 18
<211> 1098
<212> DNA
<213> Homo sapien

<400> 18
ggccgaccaa ttttttttttt ttttttttttt ttttttttttt ttctgcaagc tgctttattt 60

tttattttca tttacattag aaaataatct ctcccttgct tgattttaca agggtaaggg 120

tggtcacatg actgacagag acaaccatgg tgacacagct cttttcagct gttcatcacc 180

agcaacctgg atttcctatg cccagaacag caatgcactg aactcaagta caaattaaat 240

ttaatcccaa ctttagtcca gtctgagatt agcgcattca aagaatctgt cataacgttt 300

actatagact cttgtcgccc acagaatcag tttccagttc gtgtgtgaca tgttctattg 360

ttgaatcagt acagagttct tatgtgtgtg agttttctat ggtgactaca caaaacctca 420

ggcttacaat tgtggaggtc agaggtcaag gtgctggcag ggcaggatcc ttcctttcct 480

ccatcatggg ggctgctggc agaattcagt ttcttgcagg gctgggacgg aggtccccag 540

tcccagctgc ttaggggcca ccacactcct cggccctcct ctaaggccag cagcgcaggt 600

gcggccctcc tcgggttcta acctctcctg cttctggcat ctctcagact cagcaggaaa 660

ggctctcaag ctttaagggc ccatggggct gccctgggcc tgcaagatga cctaggacaa 720

tctccccatg tgaggcactc acaaggtctg ggggtcacaa cacgggcatc ttgggggcca 780

ttatcctgcc tacctcaccg taattccagg gtccttgaca tttttcgtaa taaaaagttt 840

aaaagtggta attacagaac tataaagctg catcggatgc cccagcccca tcaccctcca 900

gggccattcc cctcacacct gccctcccct gcagcactga gcgaatccca gacactgcag 960

agccttttcc agttcacgtc tctggaagag cccataaaac agaaacagta taaaccatag 1020

tgccattcat tatcttaccc agaagtttaa cggtcatatt ttaacatcaa atagggacta 1080

agtgttctga gtccctgg 1098

<210> 19
<211> 319
<212> DNA
<213> Homo sapien

<400> 19
agtagatcca tggggccgtg tcccagatct gccgagcggc gcagtgtgat ggattttcta 60

aagtgggggа agaaagttta tagactttcc aagcacattt atggtttttt attactatta 120

ttatggtttt aaaaagagta actttatttc tttttgtaag gaattaagta atatccttta 180

caggttctgt gaaaggactt atttttaac tgtaatattt attagtttta aaatatttgt 240

atctcatttg taacaatttg ttttaatttt ttatatatat gtttttattt ttaaaaaaca 300

taccagttga atggggtta 319

<210> 20
<211> 687
<212> DNA
<213> Homo sapien

<400> 20
atggctgagg agatggagtc gtcgctcgag gcaagctttt cgtccagcgg ggcagtgtca 60

ggggcctcag ggtttttgcc tcctgcccgc tcccgcatct tcaagataat cgtgatcggc 120

gactccaatg tgggcaagac atgcctgacc taccgcttct gcgctggccg cttccccgac 180

cgcaccgagg ccacgatagg ggtggatttc cgagaacgag cggtggagat tgatggggag 240

cgcatcaaga tccagctatg ggacacagca ggacaagaac gattcagaaa gagcatggtt 300

cagcactact acagaaatgt acatgctgtt gtcttcgtgt atgatatgac caacatggct 360

agttttcata gcctaccatc ttggatagaa gaatgcaaac aacatttgct agccaatgat 420

ataccacgga ttcttgttgg aaataaatgt gacttgagaa gtgccataca ggtacccaca 480

gacttggcac aaaaatttgc tgacacacac agtatgcctt tgtttgaaac gtctgctaaa 540

aaccccaatg ataatgacca tgtggaagct atatttatga ccttggctca taagcttaag 600

agccacaaac cattaatgct tagtcagccc cctgataatg gaattatcct gaagcctgaa 660

ccaaagcctg caatgacgtg ctggtgc 687

<210> 21
<211> 159
<212> DNA
<213> Homo sapien

<400> 21
gtcctaatca tgcgagcggc gcagtgtgat ggatgaatgt ttttaaaata tataatagga 60

cacaaagcgg cagggttttt tttgggggga gggggttgtt ttccaactca agatggcaca 120

ttagtggcca gcaatatttt ttaactcatt ccaaccagg 159

<210> 22
<211> 2687
<212> DNA
<213> Homo sapien

<400> 22
ctgaagtgca ggagacgctg gacccaattc tctctgctgg gtagttacct tatagcattt 60

ggggatttgg gttagatgat ctaaccagga ggccatcact ggatggtcac ccccccaaaa 120

aaattccatt tgagcatcaa aacctgcttt gcacaatcct atttgatgcc cccagttcag 180

cagagtcagt ggccaaagaa aactttggac gtgagtaaca cccttcagca gtcgcaacgt 240

tattttggtt ttgtgaagga ctctgaaacc atctaccctg tataaattct ggctttagaa 300

atttgcccaa gaatgctcat tctgagagct ttcctcagca gcatatatca tcagcctcat 360

cctaaaatag gcagggagcc cctcccatga gtttatccaa gttctcagct cctaaaatgc 420

aggctgccaa gaccctacac ctgccctggc tctacagcca cttacctggt ttctggactg 480

tcaccctccc agctgacctg cccgtagcca aggaatgagg acctaacttg agttggccca 540

aagtctgacc tggctgtatg tccctgtggc ccacacccag cctgtcttgc tcattcatgc 600

agcctcaaca ctggcctcca aagttccctt aacacttgca aagtcctttt tacctgtgca 660

tttggacttg aggacactgg tttctatcac aggtgagagc catgttcaat acctccagca 720

agctctcctg gctccctgca ctgtgcacgc tcctcttccc aaggtcccaa taccagcacc 780

tctagttaga gttagggtca gggtcaggcc tctcccaaca tcccagtagt ttctcctctg 840

agacacatgg gcaagagaca atttggagtc aagattttcc atttggatct attttaaatc 900

ttttagaaat gcatttgaaa cagtgtgttt gtttttttccc ttctagttaa gggactattt 960

atatgtgtat aggaaagctg tctctttttt tgtttttcct ttaacaaggt ccaaagaaag 1020

atgcaaaagg agatcacacc cttgccccgc tgagccccgt gataacaagt cactccagac 1080

taacctgtgt gccagacatt tgtgcattgt tgcactttga ggttattatt tatcaagttc 1140

ttgaaggaag cagaaagagg gactcctctc tccctccgtg tatagtctct atgtttgtgc 1200

tagtttttct ttttttctc tgtgtccagt cagccacagg gcccgcctcc ctgcaggaat 1260

aaggggtaaa acgttaggtg ttgtttggca agaaaccaca ctgactgatg aggggtaaaa 1320

tggaaccagg tagagccact ccgggcagct gtcacccatt cagaacttct ttccgcagct 1380

gaagaaatgt tcagtaacct gtttgacgct aattaaaaca gagcctgcag gaagtggggc 1440

taaagtggca ttcagtgatc ctgttctgta gacttttctt tctttttta accaaatcca 1500

aaggatgtta cagaaaagct agccactggt attttgtttt gtttaaaaaa aaaaaaaaaa 1560

aaagaaagaa agaaaaacgg aaaggaacct agctgcctgt atctttcatt tttaaaatag 1620

cacttgagtt attttctgag taatccaata aagaactttt gatgacagcc agaatgtgtt 1680

agaactctgg ctgaacattt catctcctgt gagtcagaag ggctttattt ctccctttga 1740

tggggccccc tcttctttct ggtgctctgg aagttgttta gaggaaagaa ttctaatttt 1800

aattaattgc gcagtgagtt aatctcactc gcttttctgc ttccaggcat cttaggaaaa 1860

acaaatggtt ttagtagata agggatgcct actaatgctt ttttaaaaca aacagggaca 1920

tttttattat agatttgatt tttttaatga atgtttttaa aaatatataa ataggacacc 1980

aaagcggcag ggtttttttt ggggggaggg ggtttgtttt ccaactcaag atggcacatt 2040

agtggccagc aatatttttt aactcattcc aaccaggaag ctttttttata cattgcctaa 2100

atctacgcca accagaaaat agtctcatct ctttttttct caaatgagat ccgtgtttta 2160

ttttagcatt aaattagtta cactgtgatg actggcctat tacctgactc agctccctct 2220

accttgaaat tgacattttt aaaaaatgca actaagtggt taatagtgtg tgacgctcaa 2280

agttaatgta aactggaaag gttgtgtgtc gttgcttttt gtgttttggt taggcttggt 2340

tttgtttttt aatttttata ctttctaata aatttgcagt ttcattcaaa aaaaaaaaaa 2400

aaaaaaaaaa aaaaacattt ttgggggggc ttgggcctcg gaaaaagttt ttaacaccac 2460

ttcgggtggg gcggcggggc ccacgtaggt acggcgacca cgcgggccca aacgggaccc 2520

cagaaggaaa ccctggccaa gaaaaaggtg gcgagaattc tccacaccag aaaaaaacgc 2580

gccgggggaa accgcagagt gttgcgtaaa ccacacccga agagagaact cagaagcaca 2640

caagcgggac tcaaccagga ggacccaagg gaacccgata gagtacg 2687

<210> 23
<211> 539
<212> DNA
<213> Homo sapien

<400> 23
actaaagagc acagctgctc aaagtaaagc ctgagcagtg ttctcagtaa tgtatttgaa 60

ggaaaaatac cctgatttga aaccaacagc agatgttgca aactttcata ccactgctgg 120

ccatggaagc ctcttaacaa cacactgtca tttaaggctg tgcttgtgct ttatacaaag 180

agaaagaggt ggtcttaagg ggatgcttcc agggggggtga gttcatgcct ctcctgtatt 240

ttccagcaag tggggtataa gtggtggttt gtttttaga ggggcataat aatccaggat 300

tctaagcata tggctcagct attttaaaga ggaaattaaa tattataaaa gaaatagtaa 360

agataagtta tcctcactta ggcaaaagca caggtccttt ccatatcaag tttagcctac 420

cagggttgtt ttttgtttta accctgctta ataatgttgg tgttttagaa gtagatacag 480

gcactgctct gaaaacctgg ctagccaagg atattctcag aatgttatca cctgtttgt 539

<210> 24
<211> 3262
<212> DNA
<213> Homo sapien

<400> 24
atccaacaac aatactgaga tgatctaaga aggttataac aaaatgctct tcagaaatac 60

ctaagtgctg agaattttta gtactaaaga gcacagctgc tcaaagtaaa gcctgagcag 120

tgttctcagt aatgtatttg aaggaaaaat accctgattt gaaaccaaca gcagatgttg 180

caaactttca taccactgct ggccatggaa gcctcttaac aacacactgt catttaaggc 240

tgtgcttgtg ctttatacaa agagaaagag gtggtcttaa ggggatgctt ccaggggggt 300

gagttcatgc ctctcctgta ttttccagca agtggggtat gtgtggtggt ttgtttttta 360

gaggggcata ataatccagg attctaagca tatgctcagc tattttaaag aggaaattaa 420

atattataaa agaaatagta aagataagtt atcctcactt aggcaaaagc acaggtcctt 480

tccatatcaa gtttagccta ccagggttgt tttttgtttt aaccctgctt aataatgttg 540

gtgttttaga agtagataca ggcactgctc tgaaaacctg gctagccaag gatattctca 600

gaatgttatc acctgtttgt caaagcttgt ttaaattata aaacactttt aattatatat 660

atgaggcaaa agaactaaga ctttttttcaa actaaattag aaaggagtgt cattatttga 720

ctgttaaacc aaaatatttt tggtgggtct ttttatggaa gtttaaagaa aggacatcat 780

catagatatg atctaacagt atttctaact atatttgatc attaaaagcc tcttggaatt 840

tgaagcgtga cgtgtttcta atgccccttg agaggtgaaa aataccacat aatgatcagt 900

atgctgtgcc agcttcattt ggggagaaat aactagtaga aagttctggg tgtgaggtgt 960

acagcagtct aggtggcata gtgatgaaga aagggatcag agtctgactg tcactcagaa 1020

tcctgggctc agttgcttga caaccttggg aaaattgttt tatctttgtg cgtctgtttg 1080

ctgatcttca gcgtgggaat aataacagta cctacttgaa aggatcattg tgcggattaa 1140

aagaaataat atatgtaaag cactttaaca cagcaccagg cccacggaaa gtggctaatg 1200

ttagctacta tgaatggtgc cagtgaagac actgaaaaat aagtgatttc agtaaccttc 1260

tggaaagcta tcagtttcaa ataatatttt ctctgtagta tgagatgaaa ttaaaagtgg 1320

atagctttca ggaaagataa agagaacatg cttagaatgt aagctaaaca gatttttttct 1380

gttgctcttt gaaaactatg agccctggcc agcttaacct ggtctgaggt gagactaaac 1440

acaaaaacag tagataaatc tctccctaaa agatggattc ccccacatac ccatgctact 1500

agtttctctg tctattcaca catatgtaca aatacatgaa cacagcctgt ctgtgctcag 1560

acatagagaa gtactacctg acttgagtca atgcacccaa gaagaaaagc ttggagtaga 1620

gcagaaggga gggcttggga ctcctgtctt tccagcatgc cctggggtgc agtggtcagc 1680

cacctgaaga gagagccaat agcatggggt ttacaaggca aagatagtca ttcattcaac 1740

acatattcat agagctcctt ctctgtgcca gacactgttc tggaagatag ctagatgaaa 1800

atctttgcac tcacagagct tacatgccag tgagtgaaga tcgatgataa ataaagcaaa 1860

tgcatcatat gttcacattt gataagtata tgccaaaaaa tgaagccggg aaggaggaca 1920

aggcccatgg gtgggtgttg aggtttttaa agtgtggtca ggaaaggccc cactgataag 1980

gtaacatttg agcaagtctg aaaaaggcaa ggggatcttt ggggctaact tcgggatccc 2040

tgcactttat gtaagaatgt aaacctggag tctcatttaa gaatgatcag caatacgttt 2100

agaacatatg aactgaatga aatggacatt ttttcttaat ttacgtataa atccatatga 2160

ttatacataa agttctgatg cattaataaa agcagccaaa tagggccaaa gagaaaaata 2220

acaggactct gtactggacc taactttatc attaattagg taatattttc ctcatttctt 2280

tactgctgcc attttcctca ccagtattcc agagatggtc atagctcatt actctaccac 2340

caagaaccta aaaggaatta gaatacagca gaattggcct cagtgaagag cttaaaattg 2400

ttctcctcgt agaactggac tattgatcat taccacgtga cgttggctct attactttct 2460

gttcccaatg tccttctagt ggtttgaaaa tgttaaaaca tccctaaaat ctaaatcata 2520

taatcagaat tctatagtgt cccactctat ctgtaaagat catttggaag actttagact 2580

ctattaattt taaaaggaat atttattagc catatgcaga atttctaatg atgatattgt 2640

acagcttcta attcactttt cagatcagtg tttgaaatgg caattatcag tgttggattt 2700

agttccaact acttgattta caaaaatgta catttagaga aggttaaaag aaacagtgag 2760

aaatgtaaac attcaaaatg ataattgaat ctctcagttg tgggaataat tatcagagac 2820

atgcaactga aaatgtctca cctttcatct ttttttctta attcataaag ttatcttgta 2880

gaatttgatg agaccctcct agtcattctc aactggggcg gtgctgtcac cgaaatggtg 2940

gtttgacagt gttggggcta gggcacattt ttggttgtca cagccaccgg gtggcattgc 3000

tgccgtgcat gattgtacat tatgaatgcc gcacgtgtgc tcagtaagtc tccctccaag 3060

gccgcccggg gtcagccgta tccagacttg gagcacgtgg cggtacctgt gtcgggtctg 3120

acccctggcc atgtgaactc gttctcacaa aaaaaggggg caataccggg cactctcctt 3180

ttaagccatg agttaaaacg gggaatagaa aagtttaacc ttgttgaccc actacttttg 3240

ttctcgtata taaacaacat ct 3262

<210> 25
<211> 703
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (225)..(225)
<223> a, c, g or t

<400> 25
ggtcgcggcc gaaggtcaaa ctcatggcac tgtttaccaa agagagttca ttttactgtg 60

tctaaattcg acttcaataa gagcagatta caaaatgata ttcaagagga atccagtgtg 120

tgtgtgtgcg tgtgtgtgtg tgttgtgtgt gtgtgttgtg tgtgtgtgtg tgtgctacat 180

ataataaata atcaggcggc cagcggcagt agtagtaatc actantcgtg atatactcct 240

aagcactgtt gggtgcgtcg acgagcagcg agcatgaatc accgtgaggg ataagatgat 300

gcgagaccac gccgtggaca ataagtggat gaaaccccta tctcctaaca taataaaaac 360

taacaaaata attacgacca gggctagtgg ggagctagtg tcgctcgtga taactcccga 420

gactacatca gaggagagcc gatgaggagc agaggaagaa aatcactgga tgaagccgat 480

gaggaaggga tgggaggagt aacgagatga ggccgagtaa tcacgaccaa taacatctcg 540

cagcccgtag tgataagtag agcagagaat taccacgtcg caaaaaaaaa aaaaaaaaaa 600

aaaaaagagg cgggaggaaa agaggggaaa aaagaaggac accgggggaa aaagggtaac 660

ccagggaaaa aatcccaaaa ataccacgca aaaacgaaga agg 703

<210> 26
<211> 811
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (333)..(333)
<223> a, c, g or t

<400> 26
acaaaacaaa acaaaaaaaa gagatctacc tttagtgaca cagaaatatg tttataatgt 60

acagcaaagt atatacgata agactacaga ccataggagc aatgattcaa ctgtatgcat 120

ttgtcaaact catggcactg tttaccaaag agagttcatt ttactgtgtc taaattcgac 180

ttcaataaga gcagattaca aaatgatatt caagaggaat ccagtgtgtg tgtgtgcgtg 240

tgtgtgtgtg ttgtgtgtgt gtgttgtgtg tgtgtgtgtg tgctacatat aataaataat 300

caggcggcca gcggcagtag tagtaatcac tantcgtgat atactcctaa gcactgttgg 360

gtgcgtcgac gagcagcgag catgaatcac cgtgagggat aagatgatgc gagaccacgc 420

cgtggacaat aagtggatga aacccctatc tcctaacata ataaaaacta acaaaataat 480

tacgaccagg gctagtgggg agctagtgtc gctcgtgata actcccgaga ctacatcaga 540

ggagagccga tgaggagcag aggaagaaaa tcactggatg aagccgatga ggaagggatg 600

ggaggagtaa cgagatgagg ccgagtaatc acgaccaata acatctcgca gcccgtagtg 660

ataagtagag cagagaatta ccacgtcgca aaaaaaaaaa aaaaaaaaaa aaaagaggcg 720

ggaggaaaag aggggaaaaa agaaggacac cgggggaaaa agggtaaccc agggaaaaaa 780

tcccaaaaat accacgcaaa aacgaagaag g 811

<210> 27
<211> 652
<212> DNA
<213> Homo sapien

<400> 27
agaatgataa ctcatatggg cgaatgggcc tctgatgcat gtcgagcggc gcagtgtgat 60

ggattggtcg cggccgaggt acttctaccc gagcacagac tgtgtggact ttgccccctc 120

agcagccgcc accagtgatt tctataagag ggaaacaaac tgtgacatct gctatagtta 180

atagaaatta cagtaattca gaacatggca tgggtatatc tatttttcta ccacgtctag 240

atgacactgc aaaatatgca acttggtaac acaatatccc aagcacagtt tacatgtcac 300

tatttccaat tttctgatgc taagcattca tatgaagtcc tcagacccgg tcacagcgcc 360

actcctactt tgtatgctca tagtttaaat ttttgtagga aactttcaat tgttttactt 420

tttgtataac gaacaaatgc tgtctccttt tttactaata aataatttgt attacaaaaa 480

aaaaaaaaaa aaaaaaaaaa ggcgggggggg taatcagggg ccaatacgcg ggttcccggg 540

gggagaatgg gttacccggt cacagttcca cacatttgcg agacaacaga cgggagaaga 600

ggcaggacca agacgcgagg cacgccaaga gcaagcgcac agagaaacgg ag 652

<210> 28
<211> 1511
<212> DNA
<213> Homo sapien

<400> 28
agcggagggg ggaagaaggg gagagtagga gcgggggcga aggagggagg agggcaagat 60

ggagcgcgga aaaggcggag aaaaggggcg agggagagcg ggcagaaggc aaagacagaa 120

gggagcgagg gagggagttc ctcgggcctg gccccctttac taggtcagtc tggcaggtac 180

ctcgccggcc caggacgggg ctggccaaac ctcaccgctt gctcccgggc tggcttccag 240

accaagggca cgcagaggtc ggagcctgcc cagaagccac acctggccag aaaaaccgaa 300

ggtgtatcaa ggtgtccgag tgaagatcac agtgaaggag ctgctgcagc aaagacgggc 360

acaccaggcg gcctccgggg gaacccggtc cggaggcagc agtgtccacc tttcagaccc 420

agttgcacca tcttctgcag gactgtattt tgagcctgaa ccaatttctt ccacgcccaa 480

ttatttgcaa cggggagaat tttccagttg tgtttcatgt gaagaaaact caagctgcct 540

cgaccagatc tttgattcct accttcagac agagatgcac ccggagcctt tgctcaattc 600

cacacaaagt gctccacacc atttcccaga cagcttccag gccacccctt tctgctttaa 660

ccagagcctg atcccaggat caccttcaaa ttcctccatt ctctctggct ccttagacta 720

cagttactcg ccagtgcagc tgccttcata tgctccagag aattacaatt cccctgcttc 780

tctggacacc agaacctgtg gctacccccc agaagaccat tcctaccaac acttgtcctc 840

acacgcccag tacagctgct tctcctcggc caccacctcc atctgctact gcgcatcgtg 900

tgaggcagag gacttggatg ctctccaggc ggcagagtac ttctacccga gcacagactg 960

tgtggacttt gccccctcag cagccgccac cagtgatttc tataagaggg aaacaaactg 1020

tgacatctgc tatagttaat agaaattaca gtaattcaga acatggcatg ggtatatcta 1080

tttttctacc acgtctagat gacactgcaa aatatgcaac ttggtaacac aatatcccaa 1140

gcacagttta catgtcacta tttccaattt tctgatgcta agcattcata tgaagtcctc 1200

agacccggtc acagcgccac tcctactttg tatgctcata gtttaaattt ttgtaggaaa 1260

ctttcaattg ttttactttt tgtataacga acaaatgctg tctccttttt tactaataaa 1320

taattttgta ttactaaaaa aaaaaaaaaa aaaaaaattg gcgggggggt aatcaggggc 1380

caatacgcgg gttcccgggg ggagaatggg ttacccggtc acagttccac acatttgcga 1440

gacaacagac gggagaagag gcaggaccaa gacgcgaggc acgccaagag caagcgcaca 1500

gagaaacgga g 1511

<210> 29
<211> 337
<212> DNA
<213> Homo sapien

<400> 29
gatcgactca tatgggcgaa tgggtcacat agatgcatgt cgagcggcgc agtgtgatgg 60

atgcatggtc gcggcgaggt gcaggaaaat atacagatat taaagatcag atttaattct 120

ttggtataag catgaaactg ttactgatag ctttccatgg cgagcataaa ccatgaagca 180

actcaagaag catgagagac aacaatgaaa tctagtatac aatgcagggc aggccaagaa 240

cgatgtctgc tttacaggaa aagtcaacac taacaatcta ctcctgagaa actaacacct 300

atttagatgt ttttaacata atggcaaact aaaatgt 337

<210> 30
<211> 954
<212> DNA
<213> Homo sapien

<400> 30
atgaaccggt ttggtacccg gttggtggga gccacggcga cttcttcgcc gccgccgaag 60

gcccgcagca atgaaaacct cgacaaaata gatatgtctt tggatgatat catcaagttg 120

aatcgaaagg aagggaagaa gcagaatttt ccaagactaa atagaagact cctccagcaa 180

agtggtgccc agcaattcag gatgagagtg cgatggggaa tccaacagaa ttctggtttt 240

ggtaagacta gtctgaatcg tagaggaaga gtaatgcctg gaaagagacg tcctaatgga 300

gttatcactg gccttgcagc taggaaaacg actggaattc gaaaaggaat tagtcctatg 360

aatcgtccac ctctaagtga caagaatata gaacaatatt ttccagtgtt aaaaaggaag 420

gcaaaccttc tgagacaaaa tgaagggcag aggaaaccag tagcagttct caagagacct 480

agccagctaa gcagaaaaaa taacattcca gctaatttta ccaggagtgg aaataaatta 540

aatcatcaga aagatactcg tcaggcaact tttcttttca gaagaggcct gaaggtgcag 600

gcccagttga atacagaaca actgctagac gatgtagtag caaagagaac tcgtcaatgg 660

cggacttcca ccacaaatgg agggattttg actgtatcta ttgacaatcc tggagcagtg 720

caatgcccag taactcagaa accacgatta actcgtactg ctgtaccttc atttttaaca 780

aagcgggagc aaagtgacgt caagaaagtt cctaaaggtg ttcccctgca gtttgacata 840

aacagtgtcg gaaaacagac aaggattacg ttgaaataac ggtttgggat cctgaaggaa 900

caaaaagccc ctttcccata caacaaaagg ggaaacccct ttgtccccgt ggga 954

<210> 31
<211> 260
<212> DNA
<213> Homo sapien

<400> 31
aaatgaccaa cgttacatga tttcaagggt tgtcctttct gtgcttttat ctgtcacgac 60

aggaaggtgt ggaaagttta tatccttaat ttgactactc ttggatatta aaatctttct 120

attaattaaa aagactttta gacaacctct taaatggaat tacactatgg aaaacagggc 180

tcccccaaaa acacctaggc agaactgaga gttctttgaa aaccattccc aataaaaact 240

aaatgaaaaa taaatataaa 260

<210> 32
<211> 1416
<212> DNA
<213> Homo sapien

<400> 32
ttttttatc tctgtaattc tttattaaaa atactgctgt acacatagag actgaaaaca 60

ggattaaaga tgaataacac aaattgggtc atgacattag aacctaacac actggtgctt 120

tttagggaag ttgttgacat ccaaatcaca gaaccaaggt caaaagcaaa atacaaaggt 180

accctcaaaa atatttacaa tgaagtaaat acactaacag aatttaaaac aggtacaaaa 240

tattgaaatg accaacgtta catgatttca agggttgtcc tttctgtgct ttttatctgt 300

cacgacagga aggtgtggaa agtttatatc cttaatttga ctactcttgg atattaaaat 360

ctttctatta attaaaaaga cttttagaca acctcttaaa tggaattaca ctatggaaaa 420

cagggctccc tcaaaaacac ctaggcagaa ctgagagttc tttgaaaacc attcccaata 480

aaaactaaat gaaaaataaa tttaaaacaa agcttaaaaa aatatgcatt acctgacacc 540

aacctttttct ggctgacaat atttattcat gaaaacatat cagctgtcta cctttaattt 600

gtggaccaat gttttgtgaa agctaaagag ggcaggggtt aaaatagggc ttgaatttct 660

cattctgtat agaccagcaa acttccctgt gcaaggcaag tttacatcac aaatccaaga 720

atgtttgcat cctaaatgct agtttgcttc agcccctagt taacctcagg acttggtttg 780

catataaaag gtagacagct gatatgtttt catgaataaa tattgtcagc cagaaaaggt 840

tggtgtcagg taatgcatat ttttttaagc tttgttttat atttattttt catttagttt 900

ttattgggaa tggttttcaa agaactctca gttctgccta ggtgtttttg ggggagccct 960

gttttccata gtgtaattcc atttaagagg ttgtctaaaa gtctttttaa ttaatagaaa 1020

gattttaata tccaagagta gtcaaattaa ggatataaac gatataaact ttccacacct 1080

tcctgtcgtg acagataaaa gcacagaaag gacaaccctt gaaatcatgt aacgttggtc 1140

atttcaatat tttgtacctg ttttaaattc tgttagtgta tttacttcat tgtaaatatt 1200

tttgagggta cctttgtatt ttgcttttga ccttggttct gtgatttgga tgtcaacaac 1260

ttccctaaaa agcaccagtg tgttaggttc taatgtcatg acccaatttg tgttattcat 1320

ctttaatcct gttttcagtc tctatgtgta cagcagtatt tttaataaag aattacagag 1380

ataaaaaaaa aaaaaaaaaa aaaaaaatat gcggtc 1416

<210> 33
<211> 302
<212> DNA
<213> Homo sapien

<400> 33
aagatttttc ttaattgcaa taaatattca gcatttttc taagtgaaaa tgaattgtgt 60

ttaccagtaa aagtatgcat tttaaaagac gtttcagatt tatgcttttt acgtgaagct 120

gctaaactaa aagtaaatgg aagaaaccaa gtctagtagg ttttttcttt tttaggtggg 180

ggtgggatgg gggaggttag ttacacttaa aatatcttct ccagagactg tatgctccta 240

tactagactg taagctcttt gagggcagtc tgtcagattt atctttgtat cttccccagc 300

gg 302

<210> 34
<211> 1344
<212> DNA
<213> Homo sapien

<400> 34
tttcactatt tttttttcta tctgaagctt agagatctag agctttggat ctttcgggta 60

tatgtcaatg gaggtattat tttataatac ttgcattgac atgaagtggg ttcatggggg 120

aaaaccatga gctgtgaaca tggtagcaaa caagcatata ttcatttcaa aactttcctt 180

gcttttagca gagagaagcc tgtatatgtt acatgtgtga ctttcagtag tttaaagaga 240

tgtttcaaaa aattgttgca tgtttttgat gcaatttggg aaattgttta cttcacaatg 300

tagtcattca taaaaaaaat tcatgaaaat actgaacata tgtttgagga tttttctttt 360

ccttttttaaa ttttttttatt ttttctgaga cggagatctg ctcttacgcc caggctagag 420

tgaagtggcg cgatcttggc ttactgcaac ctccaccccc caggttcaag cgattctcct 480

gcctcagcct ccggagtagc tgggattaca ggcgcccgcc accacgtccg gctaattttt 540

gtattttcag tagagacggg gttttgctat gttggccagg ctggtctcaa actcctgacc 600

tcaagtgatc cacctgcctc ggcctcccaa agtgttagga taacaggtgt gagccaccgt 660

gcccggctga agatttttct taattgcaat aaatattcag catttttct aatgaaaatg 720

aattttgttt accagtaaaa gtatgcattt taaaagactt tcagatttat gctttttacg 780

tgaagctgct aaactaaaag taaatggaag aaaccaagtc tagtaggttt tttctttttt 840

ttgtgggggt gggatggggg aggttagtta cacttaaaat atcttctcca gagactgtat 900

gctcctatac tagactgtaa gctctttgag ggcagtctgt cagatttatc tttgtatctt 960

ccccagcgcc tagtgtagtg ccttgcacat aataggcgcc caataaatat tgatgaagaa 1020

tgaaggcgtt gtgtttctaa tgtgaccaaa ccatggggat tctttgtcat taataccgtc 1080

ctcctttgta agtgctgttt tttttttca ttcttgagct cctaatgaca ttagatctta 1140

tcaggggcag ttggacagtt cagtaaaggt aaatgctgct cttgctctag ttgctgtgac 1200

ctatgttctt tctgacttgc taagagagcc aagtgatagt ggctagtgat aagattgata 1260

cataaattgc tttactttga aataacactg gaaaacccta ccgtagacct gatcaagaaa 1320

aaaaaaaaaa aaaaatgagc ggcc 1344

<210> 35
<211> 163
<212> DNA
<213> Homo sapien

<400> 35
gggcggccgc cgggcaggta cctataaatg tcttctgctg ctaatattta tctcagcact 60

ttctaaaccc aaaagtgcta cctaagaaga aatttagcca aaaaataccc agctaaggta 120

gccatagcca agtgtattta agtatgttat agaatatatt tga 163

<210> 36
<211> 643
<212> DNA
<213> Homo sapien

<400> 36

ttcatttccc gaactgaagt atggaaattt ggtaatgttg tcattgaaca tctataccac 60

tggatacaca tctgttcagc tctcatgaag ataaccaaac aactaaatag tggtattaca 120

cctccgttgc cctccaagac tgacaattat atgtatgcaa aaatgccagg ggaaggtttg 180

caagagaagt gataatggat gataatggaa ttgatactgt atttaggatc ctttgtttgt 240

tatcagtttt gtttgttaac tataaaatat tttccattgg aaaggggtac ctataaatgt 300

cttctgctgc taatatttat ctcagcactt tctaaaccca aaagtgctac ctaagaagaa 360

atttagccaa aaaataccca gctaaggtag ccatagccaa gtgtatttaa gtatgttata 420

gaatatattt gaaagcttcc tttcagtttg agctttgtat ctgctgtgga actgttatgg 480

ttgattgggt agttattttt cattcttata aggttcaaag taacagctga ggatttagaa 540

aacaagaata ccaaatagaa tacgaaataa taaagataaa ccaaaagaat accaaataat 600

aaagattttt aagaaatgga aaaaaaaaaa aaaaaaaaaa att 643

<210> 37
<211> 478
<212> DNA
<213> Homo sapien

<400> 37
gcgtggtcgc ggcgaggtac aaaaataaca gcatttagtt gcagattaga aacagatgtg 60

aagggcgaaa aagcaccata gggaaggaca taagaggtcc ctggagtcag acttgggaga 120

tgtgagtttt atcagttttg ccattaggta gttgtgtgca cccttgggca tatagcactt 180

ttttggtaat tctatttttcg cacttttcaa atgagatgca attagattag agactgtaaa 240

gtaaaagctg ccatgcttca ttttttttaaa accaattaaa cgccattttt atacggaagt 300

ttggacaaac aaaaacaaca aaaaaacaac aacaaaacag cttgggcggc tacttcggtg 360

gctcattacg cggtttccct ggtggtggac attgggtttc tccgctccac aattccccag 420

acaacttagg gacgcaagaa accccgatca caaaagcact cccacaacca cacacaca 478

<210> 38
<211> 833
<212> DNA
<213> Homo sapien

<400> 38
ccgggccggc cgggcaggta cactatttgc actgtatgct ggcgcgttta ctgcttatga 60

ttaaaagttt agaccctcat acgaggtttg caatggttac tttaagtagg acggagattc 120

ccctagtcct ctataaaaga taatccactt tatcgctact acgattccgt tatttataga 180

aagagaagat cgttctcgta gtacacatgt ttatggagga atatcttaag atagaacact 240

aattcatatc tatgacaaaa aaaatcacgg tagttcgcaa catcgtaccc atggcatctg 300

gacttcttgc gctaaccgta gttacctgtg tatagaatcc acgttgttaa tcaatcagtg 360

aatcttcatt ctgcgcctga ttcgagaagt agaagacccg tcttctctac tttctcggct 420

ctaaacttta ctgactcaaa cgaagaagct gggcaactga caaaacagga caggttgttt 480

ttaatccagt ctacaaataa acaagacaat gcctgagtta gcctctata tagatttcag 540

gcttatgctg acctcgtggt aaaatctgta tttaactaaa agttaataaa aatacatatt 600

gttcatttta aaataattac tgattttgct tggggtaatc ccaacccctt accccaaatc 660

atatattttt aggacaagat ttcctgcata accacaacct ggttcctcca cccacccatc 720

atagatgttt caataagaac cctggatcag gagaagcatc tctatctaca tgcttgtctg 780

ctaggaggct aaagcttggg taacatgcca gctggtctgg tgaatgttcg tca 833

<210> 39
<211> 718
<212> DNA
<213> Homo sapien

<400> 39
gccgggcagg tacttttta aatgttaaaa atactagagc tgtattaact tcgtgatttt 60

atttttcttc ttagcactaa cttcaaaata accatacagt acagttttta aaatttacat 120

tcacagagaa ttttaatgac attggaaaat gtaagaaatt tgaaaaaaag atggagtaaa 180

atatgtataa aattgataat agttgattta gggtggtaga agtaaacata atttttctg 240

tttatatttt tctctatctt ttaaattttg ctaatgtgca tagattcttt taaaataata 300

agaaaataat aaagttaata cgttataaaa aatagggacc tggctgttga agtgcgatgg 360

agacaatttg ttagaacatg tggcttgtta cacagacgct tgagaagttg ttgagagaga 420

acgattacct agaaacaaga gttacagtaa atggggtaaa aagggcaaaa gttcttcaga 480

ttactatcct atttaccaaa gtttgtgata tgtatttctg aatatattgt tgaagagctt 540

cacttctatc aagccatagc acttatttgt cactctgata taacaattta acataaaaac 600

cactcccaaa cagttaaaac cagctctaat ttccaatctg cagagtttta agcaaatgcc 660

ggattgtctg gacagagaaa atcctccaga ggagagccag agaaaataga tgtgaggg 718

<210> 40
<211> 1439
<212> DNA
<213> Homo sapien

<400> 40

gccgcaattt ttttttttt ttttttttt ttttttctgg acacaatatg tttaatatta 60

gaagaatgat tacacatagc ttgttacaga tttccaaaaa acagtaggta cagtttttaa 120

aatttacatt cacagagaat tttaatgaca ttggaaaatg taagaaactt tgaaaaaaga 180

tggagtaaaa tatgtataaa attgataata gttgatttag ggtggtagaa gtaaacataa 240

ttttttctgt ttatattttt ctctatcttt taaattttgc taatgtgcat agattctttt 300

aaaataataa gaaaataata aggttaatac gttataaaaa atagggacct ggctgttgaa 360

gtgcgatgga gacaatttgt tagaacatgt ggcttgttac acagacgctt gagaagtttg 420

ttgagagaga acgattacct agaaacaaga gttacagtaa atggggtaaa aagggcaaaa 480

gttcttcaga ttactatcct atttaccaaa gtttgtgata tgtattttga atatatgtga 540

agagcttcac ttctatcaag ccatagcact tatttgtcac tctgatataa caatttaaca 600

taaaattgag ttcattcaaa tgagcagaaa aggaaaaaaa tgtaagtatg tctactttcc 660

cgggaatggt cttgcaccag tatctttcta ttcatgttag cattttctat gtaagaaaca 720

aatacccaaa gacttttgta gtagagactc catctgttcc aatatagtca atatccttct 780

atttgagcat caattagtgg ccttcaatta accaccttgc attcggtaat agtctgaagg 840

ggagagttct tgattctggg aatcaaagag ctttactgct gtgcctcatg cagagagcag 900

accagatgtc ttctaaaagc gaggcagtct cctttaaata tgcattagag ctagcattac 960

tatcacactt agccttccaa ggctctaaaa gcagtggcaa aggagggcta aacatacaaa 1020

atgcaaacaa cttggtctgt aagcagtcag tatgtcatta tccttcaaca gaactctttc 1080

aattgaatgt ttgtggttta gaggttttag gatataatat ttctcacttg aaagagtttt 1140

tttatattac tatatgaagc catggtgcat ttaactgact taataaaatg taattcttac 1200

tttaagtctt gagaggagaa aagcctctgt gaaagaaatc tttgttagca aggcatataa 1260

gcagagtcct ggtctgcaat aatattgatg atcacgactt gtgtgttact atataaaatt 1320

caaccagtca aaattcaaca tctttaagaa tattgctact ttgggcaaaa tttgagtttc 1380

attagagtaa aatcatttct gacatttcat aaagtttaat gcaaacaaaa atgattaat 1439

<210> 41
<211> 298
<212> DNA
<213> Homo sapien

<400> 41
gcgggcaggt aactgctgag attaagacaa ttgtggatgt gtatgtctag gtttgaatct 60

ctgggctgca gatttgtttt gccctggcag agaaagagga gtctttgggg aggtgagctg 120

tttcttgtga tttcaggcaa gaggcacata gaaactttgt atgagtgggg attttgtttt 180

aagtgctgga aaattagggc aggaattacg tgtttgcaag ttgtgccatc actggtttga 240

atttgactgc ctcatcaagg ggcaagagtt attcttgaag atctcattct cccagaaa 298

<210> 42
<211> 2023
<212> DNA
<213> Homo sapien

<400> 42
gggtttttttc tttgtttcaa gacaggaagc agtctggtta agggagaact agtggaaagg 60

gttaaatgac aggttaagtt gagtacaaag ctttcccagt actgctgaga ttaagacaat 120

tgtggatgtg tatgtctagg tttgaatctc tgggctgcag attgcttttg ccctggcaga 180

gaaagaggag tctttgggca ggtgagctgt ttcttgttga tttcaggcaa gaggcacata 240

gaaactttgt attgagtggg gattttgttt taagtgctgg aaaattaggg caggaattac 300

gtgtttgcaa gttgttgcca tcactggttt gaatttgact gcctcatcaa ggggcaagag 360

ttattcttga agatctcatt ctcccagaaa cagaacttta gggaaaatgg ctgtggctta 420

gcttttcagc tgatgcaggg taataagctt tctggttggt tttccttcca attctggaaa 480

ggtgtccaca ctaagaccct taactctagg gcttgcataa gtattctagc atcgttagct 540

aatgagttgg tcattgtttc tctttatcaa taattgtgtt aataccaatc ttataattta 600

aaaattatct tgtatgtaag agaagttcgg ggttagggag gaagaggagc aaagtgggat 660

attttctctt taatgcttag atactgtttc ttccctaaga tgtgtttctc aaccacaatt 720

ggtggaatga accagagagg caagaggaag tgaattgcac caatttagtt tagcgactgt 780

gccttttgca ggaaaaactg ggtgaatcac agctcctcag agtcctggac tcaactagaa 840

ttgaagatag acttattttg ctgactgggc ttcttagagt ttatgtgact tgaacagctt 900

ggcccctgcc tcccttctgc tactgtgagc agccttcctt cttcctggaa tgcagttctc 960

ttgcttatga tcctatgaat aaggcaaaat ggctggtctt tgtaaggcag gtcttgccct 1020

agcttctcag aaacaggagc attttaggat cagtattagg agatgcccca gggagtaaga 1080

aagtattggg ttcagtgata aatctggact ctgacacttc ttttactctc cctctttaat 1140

actaaaagct ctgcataagc aatggttcag aacctgtctt gggtacagac ctgttgaatc 1200

tgacagaaac cagaaatgca cttttgagaa aaagacattt gtaattcact cagttttcca 1260

tacacattta gcaggttcaa agcccatctg tggaatccct aaactgcctt caaagaaagg 1320

gagttccccg atctaaaatg gtcattatat atttgtgtca agaattagaa ggcaagggtc 1380

actaaatatt ttaaggatta aggtaccaga ggcatcagtg tataaggatg gagtctggtc 1440

tttaattacg acaagggtat tgcttacatt ctactctctg gttttcaaaa agatctgaca 1500

tgctgacaaa tccagctcct cacaaatctt gtttgaagga cttgtgggaa gtgatattcc 1560

ttactattag atcacgcccc ttataactac atgttaacat ccagcctttt atctgtttga 1620

gtaattgtag ggatagaaag tgaagccccc agagttaggt gcaagtatag cacccagctg 1680

aaaggcatca tggagtctaa gggccttcta cagaaggggc aatcctttgg gttatttctg 1740

gtgtaccact gtcttctcta cctcggtcca acaccacctc tcttggacaa aaaataaaac 1800

aagcaacagc catcagatga gtgaatagat ttgaatgatt tttcccacag ggaatcagcc 1860

tcaaatgttc atgtttcacc ccgtcccctt taaataaaaa gaatctctgt gttctctttg 1920

ggcaaaatgt aaaacaggga tatcatcttc aggaacctgt cacatttttc catctggtac 1980

ctccacccta ttctgagtat cctccccttt ccaccccaac ata 2023

<210> 43
<211> 667
<212> DNA
<213> Homo sapien

<400> 43
tggtcgcggc cgaggtctgg cctggggctt cctcacccac aaacaccatg cttcctgcag 60

ggacactggt gggtgctggc ctgggggttc ctcacccaca aacaccatgc ttcctgcagg 120

gacactggtg ggtgctggcc tgggggttcc tcacccacaa acaccatgct tcctgcaggg 180

acgttgatgg gcgctggcct ggacgttcct cacatacaac cgccatgctt cctgcgggga 240

cgctggtggg cgctggcctg gggcttcctc acatacaaac cccatgcttc ctgcaggggc 300

gctggtgcgc cctggcctgg ggcttcctca catacaaacc ccatgcttcc tacagggcac 360

gctggtggac tgctggccct gaggcttcct cacacacaat tgctatcctt ccgcacggca 420

cgctggctgc gcgcactggc ctggggctac ctcacccaca aaccccatgc cttcctattg 480

attaacctga gctacccgcg ctctccctga caacggtgga caaagatttc ccacacggcg 540

gcctgcgcac gtggctcaac cagaagcccg cagccctcca tggcaacgca tccttccccg 600

aaccacacat ccagcaccac ccaagaagcc gcagcaccag cccgccccag cccggccccc 660

acccccc 667

<210> 44
<211> 495
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (220)..(220)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (234)..(234)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (262)..(262)
<223> a, c, g or t

<400> 44
gcgtggtcgc ggccgaggta ccactgcact ccagcctaag caacagagta agaccctgtc 60

tctaaaagaa aaaaagaaaa agaaatagaa catttccaga tctcagaagt cttctcttgt 120

cactatccct tacaaaggca acctgacttt taataccata gattaatttt gtctgttttt 180

atactttata taaatgtaat caatcaatat gcaatctttn gtgtcagctt cttntgctct 240

acattatact tgtgagatcc anaaaaaaaa aaaaaacaaa aaaaaaaaaa acggcttggg 300

gcggtaacct caaggcggcc aataaggcgg ggtctcgcgg gtggtggaaa tatgggtgta 360

tactcgggcg ctcaaaatat cccaacacac aacactatat caagcggcac ggcaaaaaag 420

ggggaaaacc gaaaacaaga aaacagaaaa aaaaaaagaa aaaaaaaaaa aaacagaaaa 480

aaaaaaaaaa acgaa 495

<210> 45
<211> 651
<212> DNA
<213> Homo sapien

<400> 45
cggccgccgg gcaggtacta atttccattc tcaccaacag ttcactaggg ttccctttc 60

tccacattgt tgccaacatt cttaatcttg tgtttttaa taacagctat cctaacaggt 120

atgaggtgat ctctctcatt gcggttttga ttcgcatttc cctaacggtt ggtgatactg 180

agcatttttg catacaccgg gtcatttgtt ctttgttgtt gacttgagat cccttatata 240

gtttggatac tgctgtggcc tgaatgtttg tgtcccccaa aaattcgtat attgaactct 300

catccctaag gtcaacagtt tagggaagcg attaggtcct gaggactctg ccctcttgca 360

tagaattagt gctcttataa aagatgcccg agggagctct tttgcccctc ctgccatgtg 420

aggacacagc tagaagctac catctgtgaa ccaggaagcc cccctcacca gacactgaat 480

ctgctggagc caccatcttg gacttcccag cctccagagc tgtgagaaat acatgcctgt 540

agttaagcaa aaaaaaaaaa aaaacaacaa aaacagcgtg ggggaaacaa ggacaaaaga 600

ggtcacctgg gtaaaaggga actcggacca cattccaaca cttacacaaa g 651

<210> 46
<211> 873
<212> DNA
<213> Homo sapien

<400> 46
atgctgcgcc gcgaggcccg cctgcgccgc gagtacctgt accgcaaggc ccgggaggag 60

gcgcagcgct cagcccagga gaggaaggag cggctgcggc gcgcgctgga agaaaaccgc 120

ctgattccca ctgagttacg ccgagaggct ctggccttac aggggtccct ggagtttgat 180

gatgctggag gtgaaggtgt gaccagccac gtggatgatg aataccgatg ggcaggagtc 240

gaggatccca aggttatgat cactacctcc cgagacccca gttcccgcct caagatgttt 300

gcaaaggagc tgaagctggt gttcccgggc gcccagcgaa tgaaccgagg tcgacatgaa 360

gtgggggcac tggtgcgagc ctgcaaagcc aacggcgtca ccgatctgct ggtcgttcac 420

gagcatcggg gcacacctgt ggggctcatc gtcagccacc tgccctttgg tcctactgcc 480

tacttcacgc tgtgcaatgt ggtcatgcgg catgacatcc cagacctggg caccatgtcg 540

gaggccaagc cccacctcat cacacacggc ttctcctccc gcctgggcaa gcgggtctct 600

gacatcctcc gatacctatt tcccgtgccc aaagatgaca gccaccgggt catcaccttc 660

gcaaaccagg acgactacat atcattccgg caccatgtgt ataagaagac agaccaccgc 720

aacgtggagc tcactgaggt cgggccccgc tttgagctga agctgtacat gatccgtctg 780

ggcacgctgg agcaggaggc cacagcagac gtggagtggc gctggcaccc ttacaccaat 840

accgcacgca agagagtctt cctgagcacc gag 873

<210> 47
<211> 213
<212> DNA
<213> Homo sapien

<400> 47
tatgagtata agggcatggt ttcctctaag ctgtcgagcg gcgcatgtga tggatccggg 60

caggtactgg acacctggca tgctgactgc cacgtgcagg caagaaacat ctgtccagta 120

agttaggggg aagacgggat ggggaataaa ccctcggaaa tctctgcaca ccactcttgg 180

tgctatgctt ttaattctgt ttccctttct cct 213

<210> 48
<211> 658
<212> DNA
<213> Homo sapien

<400> 48
ggcgaaaccg gaacagagaa tttatcactt ctgggactca cagtcgtgat gtctttcaag 60

agggaaggag acgattggag tcaactcaat gtgctcaaaa aaagaagagt cggggacctc 120

ctagccagtt acattccaga ggatgaggcg ctgatgcttc gggatggacg ctttgcttgt 180

gccatctgcc cccatcgacc ggtactggac accctggcca tgctgactgc ccaccgtgca 240

ggcaagaaac atctgtccag taagttaggg ggaagacggg atggggaata aaccctcgaa 300

atctctgcac accactcttg gtgctatgct tttaattctg tttccctttc tcctcaggct 360

tgcagctttt ctatggcaag aagcagccgg gaaaggaaag aaagcagaat ccaaaacatc 420

agaatgaatt gagaagggaa gaaaccaaag ctgaggctcc tctgctaact cagacacgac 480

ttatcaccca gagtgctctg cacagagctc cccactataa cagttgctgc cgccggaagt 540

acaggtatgg gacgggaaag ccagaggtag gaaggctcag aaggagacag atggctctaa 600

aagagttttc cagtgtgtat tctgaggaat actagtgttc tggagatgtt acttagtg 658

<210> 49
<211> 703
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (169)..(169)
<223> a, c, g or t

<400> 49
ccgaggtaca ttcaaacagt tatacaacta tcaccactat tccaattcca gaacattctc 60

atcatcgccc aaagaaacca catacctatt agcagtcact ccccatcctc cctttctcag 120

cccctggcaa ccactccctt aagtgaagag tgacaacttt cctgggcant gtgctttcag 180

tagtatgtgg ctttacatgt ttccattaga atttttaaca ccaaattcaa gcagtgagct 240

tgtaactatt ctgagattat gaaatatcct tttatataca actatttttg tctcaaacat 300

gtttctttat acataaaaaa tagatatttc tgtttccatt ttttaatcaa attctgtcct 360

tatttcagaa gtgagaaaaa tcaatactcc aatattaaaa agcaggaata accatagttc 420

tattattaac tgtgggccac cacactctct gtcctactgc ttcccacaga atctgaggtg 480

ccaagggctg caaggccttt gagggcaagc tgcacatttt acagatgaag aaacagatcc 540

gacatgggct tgtgacatgt ccaaggtcac aaggccagtt aacagcaagc taggatgaga 600

atccttctta ctagaactta gtattaatat taatgcgaca gctgggtatc atgtcatagc 660

tgttccggtg aatgtatcgt caaaaaaaaa aaaaaaaaaa aaa 703

<210> 50
<211> 1251
<212> DNA
<213> Homo sapien

<400> 50
aaaaaggccc tgagtggaac tgtattatcc agaagtaagc tagtttttac atggaggatt 60

atgcagttta cataattgaa atgtgttttt ctctgtgtgc tgttctcata ttccaatatt 120

ctttttcct ctcatggtca tgatgttttc ttttgagata taattcacat accataaaat 180

tgatgctttt aaactataca attcgttagc tgggtgtggc agcacacacc tgtagtccca 240

gctactcagg aggctgaagt gagaggatca cttgaactgg gaggcagagg ttgcagtgaa 300

ccgagattgc gccgctgcac tccatcctag gcgacagggt gagcccctgt ctcaaaaata 360

aataaataaa caattcagtg gttcctagta cattcaaaca gttatacaac tatcaccact 420

attccaattc cagaacattc tcatcatcgc ccaaagaaac cacataccta ttagcagtca 480

ctccccatcc tccctttctc agcccctggc aaccactccc ttaagtgaag agtgacaact 540

ttcctgggca ttgtgctttc agtagtatgt ggctttacat gtttccatta gaatttttaa 600

caccaaattc aagcagtgag ctttgtaact attctgagat tatgaaatat ccttttatat 660

acaactattt ttgtctcaaa catgtttctt tatacataaa aaatagatat ttctgtttcc 720

atttttaat caaattctgt ccttatttca gaagtgagaa aaatcaatac tccaatatta 780

aaaagcagga ataaccatag ttctattatt aactgtgggc caccacactc tctgtcctac 840

tgcttcccac agaatctgag gtgccaaggg ctgcaaggcc tttgagggca agctgcacat 900

tttacagatg aagaaacaga tccgacatgg gcttgtgaca tgtccaaggt cacaaggcca 960

gttaacagca aagctaggat gagaatccct tcttactaga actttagtat caaatattta 1020

aatgctgact ttgtgggtaa cctaattcag ctaccacatg aatctaatta tgtcagtttc 1080

ctctacagct ttgatctgag catgtgattt ctttttttt accattttaa aaacatttac 1140

atgttatctt ttaagacctg taaggacatg actagtctat ttagccagag ggcccaaatc 1200

actcactgag acaaaacaaa gaagagccaa agttccagag ggacctgaga g 1251

<210> 51
<211> 402

<212> DNA
<213> Homo sapien

<400> 51
cgagcggccg cccgggcagg tacccgctca gagattatcc acagcagcca gatggttcta 60

ccttccacaa agattgtggt tgcaattctg ggcttctaag ttctggttac ttcatatttt 120

tccttttgtt cctccagccc tagaggtggt agctgctttc tgaagttatt atttctagat 180

gacttttggt ttttcagcct ttgtattttg cttttcagcc ctctaatgcc tgtataacca 240

atttccctgt aactaaataa atttcctcca ttgaaaaaaa aaaaaaaaaa aaaaaaaaaa 300

ggttgtgtgg ggttattcgg tggctctagg gcgtgttccc tgtgtgtgtg gaatgtggtt 360

ttcccggtcc aaaatttccc caaaaaattg cggacacacc tg 402

<210> 52
<211> 1042
<212> DNA
<213> Homo sapien

<400> 52
caattgttct caaacttcac tagccccgtc ggcgcggacg cttgtcgaga atgcagattc 60

ctgggtactg ccagatacga attgagcata ccacaaaaaa gttctcattt tgtgtcctcc 120

catcccattc tcctcactaa ccaaaggcta ggaattatct gtgaatgtag gaccactgga 180

tttgcagtct tcatctgaca ctgtggagag tttctaggaa tgaaacagat atatggcctt 240

gggtcccctt ttttttcctt ttttttttt ttaatagaga cgagcatctc actatgttgc 300

ctagggtagt cttgaactcc tggcctcaag caatccccac ccgactccgc ctctcgaagt 360

gatgggatta caggcataaa ccaccacgcc tggccagaag gtgctttaac accaaatctg 420

aaaattgttc agaagagaaa cattgagcat gaacaccatc tgtgcgagtc atttacttat 480

tgcccctcac ctctaaatct accttctgta ctcttcttcc ctgtaatgat ggggctagtt 540

gtcctcaaac tgtttctcag acttctttt aagcttgctt cctgttcagt tctgccaata 600

ggggtcacta gagagagact gggaggcaga aggagagaat atgcttcctg ttttttctgt 660

tcttgttaat gttgcttaca ggaccagcaa tgcttcttca cctagagaca cttctcccag 720

cagtggcagt gccacttcag cttctttcag cactactgga atcagcctca gtgattcccc 780

ctgtacccgc tcagagatta tccacagcag ccagatggtt ctaccttcca caaagattgt 840

ggttgcaatt ctgggcttct aagttctggt tacttcatat ttttcctttt gttcctccag 900

ccctagaggt ggtagctgct ttctgaagtt attatttcta gatgactttt ggtttttcag 960

cctttgtatt ttgcttttca gccctctaat gcctgtataa ccaatttccc tgtaataaat 1020

caatttcctc cattgaaaaa aa 1042

<210> 53
<211> 240
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (44)..(44)
<223> a, c, g or t

<400> 53
tcattagatc atgtcgagcg gcgcatgtga tgatgcggcg ccgngcaggt ttttttttt 60

ttgaacacaa gggtcagttc ttcaattcat gagcagtcag aacaggagat gcttaggaag 120

gaatcgtggc tggtgcctct tctccatgct catcccatac cccagtgaca ggataccgtt 180

ccctgaagtt taaaaacatg caccacactt ccggtaaagg ctggagccac agaggcacct 240

<210> 54
<211> 1590
<212> DNA
<213> Homo sapien

<400> 54
atggaaagga tggtgggctc tggcctcctg tggctggcct tggtctcctg cattctgacc 60

caggcatctg cagtgcagcg aggttatgga aaccccattg aagccagttc gtatgggctg 120

gacctggact gcggagctcc tggcacccca gaggctcatg tctgttttga cccctgtcag 180

aattacaccc tcctggatga acccttccga agcacagaga actcagcagg gtcccagggg 240

tgcgataaaa acatgagcgg ctggtaccgc tttgtagggg aaggaggagt aaggatgtcg 300

gagacctgtg tccaggtgca ccgatgccag acagacgctc ccatgtggct gaatgggacc 360

caccctgccc ttggggatgg catcaccaac cacactgcct gtgcccattg gagtggcaac 420

tgctgtttct ggaaaacaga ggtgctggtg aaggcctgcc caggcgggta ccatgtgtac 480

cggttggaag gcactccctg gtgtaatctg agatactgca cagacccatc cactgtggag 540

gacaagtgtg agaaggcctg ccgccccgag gaggagtgcc ttgccctcaa cagcacctgg 600

ggctgtttct gcagacagga cctcaatagt tctgatgtcc acagtttgca gcctcagcta 660

gactgtgggc ccagggagat caaggtgaag gtggacaaat gtttgctggg aggcctgggt 720

ttgggggagg aggtcattgc ctacctgcga gacccaaact gcagcagcat cttgcagaca 780

gaggagagga actgggtatc tgtgaccagc cccgtccagg ctagtgcctg caggaacatt 840

ctggagagaa atcaaaccca tgccatctac aaaaacaccc tctccttggt caatgatttc 900

atcatcagag acaccatcct caacatcaac ttccaatgtg cctacccact ggacatgaaa 960

gtcagcctcc aagctgcctt gcagcccatt gtaagttccc tgaacgtcag tgtggacggg 1020

aatggagagt tcattgtcag gatggccctc ttccaagacc agaactacac gaatccttac 1080

gaaggggatg cagttgaact gtctgttgag tccgtgctgt atgtgggtgc catcttggaa 1140

caaggggaca cctcccggtt taacctggtg ttgaggaact gctatgccac ccccactgaa 1200

gacaaggctg accttgtgaa gtatttcatc atcagaaaca gctgctcaaa tcaacgtgat 1260

tccaccatcc acgtggagga gaatgggcag tcctcggaaa gccggttctc agttcagatg 1320

ttcatgtttg ctggacatta tgacctagtt ttcctgcatt gtgagattca tctctgtgat 1380

tctcttaatg aacagtgcca gccttcttgc tcaagaagtc aagtccgcag tgaagtaccg 1440

gccatcgacc tagcccgggt tctagatttg gggcccatca ctcggagagg tgcacagtct 1500

cccggtgtca tgaatggaac ccctagcact gcagggttcc tggtggcctg gcctatggtc 1560

ctcctgactg tcctcctggc ttggctgttc 1590

<210> 55
<211> 467
<212> DNA
<213> Homo sapien

<400> 55
gtcgcggccg aggtacttat ataagggtta tttttaaagt caggaatttt ctcaaggaaa 60

attttaagct actacaggcc aggtgcagtg gctcacacct gtaatcccag cactttggaa 120

ggccaagggg gggcggatca cgtaaggcca ggagttaaag accagcctgg ccaacatggc 180

gaaaccccgt ctccactaaa aatacaaaaa ttagctgagg gtggtggtgc atgtctgtaa 240

tcccagctac tcgggaggtg gaggttgcag tgagctgaga tcacattgct tcactccagc 300

ctgggcgaca gagtgagact gtttaaaaaa aaatttttt aagctactgc aataaatttg 360

tttattcatc aaataaaata aatagcaagg attttcttct attggaaaaa atagatagca 420

aggattttct tctagtggaa aaagtttctc ctgtttaacc tggcatt 467

<210> 56
<211> 2970
<212> DNA
<213> Homo sapien

<400> 56
atgtcggaag aaacccgaca gagcaaattg gccgcagcga agaaaaagtt gagagaatat 60

cagcagagga atagccctgg tgttcctaca ggagcgaaaa agaagaagaa aataaaaaat 120

ggcagtaacc ctgagacaac cacttctggt ggttgccact cacctgagga tacacccaag 180

gacaatgctg ctactctaca accatctgat gacaccgtgt tacctggcgg tgtcccttcc 240

cctggtgcca gtctcactag catggcggca tctcagaatc atgatgctga caatgtccct 300

aatctcatgg atgaaaccaa gactttctca tcaaccgaga gcctgcgaca actctcccaa 360

cagctcaatg gtcttgtttg tgagtctgcg acatgtgtca atggggaggg ccctgcatcg 420

tctgctaacc tgaaggatct ggagagccgg taccaacagc tagcggtagc cctggactcc 480

agctatgtaa caaacaaaca actcaatatc acgatagaga aattgaaaca acagaaccaa 540

gaaattacgg atcagttgga agaagaaaag aaagaatgcc accaaaagca gggagcccta 600

agggagcagt tacaggttca cattcagacc atagggatcc tcgtatcaga gaaagctgag 660

ttacagacag ccctggctca cactcagcat gctgccaggc agaaagaagg agagtctgaa 720

gatctggcca gccgcctgca gtattcccgg cggcgtgtgg gagagttgga gcgggctctc 780

tctgctgtct ccacgcagca gaagaaggca gacaggtaca acaaggagtt aaccaaagag 840

agagacgccc tcaggctgga gttatacaag aacacccaaa gcaatgagga cctgaagcaa 900

gagaaatcag aattggaaga gaagcttcgg gtcctagtga ctgagaaggc tggcatgcag 960

cttaacttgg aagaattgca aaagaagtta gagatgacgg aactcctgct tcaacagttt 1020

tcaagccggt gtgaagcccc tgatgctaac cagcagttac agcaggccat ggaggagcgg 1080

gcacagctgg aagcacacct ggggcaggta atggagtcgg ttagacaact acaaatggag 1140

agagataaat atgcggagaa tctcaaagga gagagcgcca tgtggcggca gaggatgcag 1200

cagatgtcag agcaggtgca cacattgaga gaggagaagg aatgtagcat gagtcgggta 1260

caggagctgg agacgagctt ggctgaactg aggaaccaga tggctgaacc cccgccccca 1320

gagcccccag cagggccctc cgaggtggag cagcagctac aagcggaggc tgagcacctg 1380

cggaaggagc tggagggtct ggcaggacag cttcaagccc aggtgcaaga caatgagggc 1440

ttgagtcgcc tgaaccggga gcaggaggag aggctgctgg agctggagcg ggcggccgag 1500

ctctgggggg agcaggcgga ggcgcgcagg caaatcctgg agaccatgca gaacgaccgc 1560

actaccatca gccgcgcact ctcccagaac cgggagctca aggagcagct ggctgagctg 1620

cagagcggat ttgtaaagct gactaatgag aacatggaga tcaccagcgc actgcagtcg 1680

gagcagcacg tcaagaggga gctgggaaag aagctgggcg agctgcagga gaagctgagc 1740

gagctgaagg aaacggtgga gctgaagagc caagaggctc aaagtctgca gcagcagcga 1800

gaccagtacc tgggacacct gcagcagtat gtggccgcct atcagcagct gacctctgag 1860

aaggaggtgc tgcataatca gctactgctg cagacccagc tcgtggacca gctgcagcag 1920

caggaagctc agggcaaagc ggtggccgag atggcccgcc aagagttgca ggaaacccag 1980

gagcgcctgg aagctgccac ccagcagaat cagcagctac gggcccagtt gagcctcatg 2040

gctcaccctg gggaaggaga tggactggac cgggaggagg aggaggatga ggaggaggag 2100

gaggaggagg cggtggcagt acctcagccc atgccaagca tcccggagga cctggagagc 2160

cgggaagcca tggtggcatt tttcaactca gctgtagcca gtgccgagga ggagcaggca 2220

aggctacgtg ggcagctgaa ggagcaaagg gtgcgctgcc ggcgcctggc tcacctgctg 2280

gcctcggccc agaaggagcc tgaggcagca gccccagccc cagggaccgg gggtgattct 2340

gtgtgtgggg agacccaccg ggccctgcag ggggccatgg agaagctgca gagccgcttt 2400

atggagctca tgcaggagaa ggcagacctg aaggagaggg tagaggaact ggaacatcgc 2460

tgcatccagc tttctggaga gacagacacc attggagagt acattgcact gtaccagagc 2520

cagagggcag tgctgaagga gcggcaccgg gagaaggagg agtacatcag caggctggcc 2580

caagacaagg aggagatgaa ggtgaagctg ctggagctgc aggagctggt cttacggctt 2640

gtgggcgacc gcaacgagtg gcatggcaga ttcctggcag ctgcccagaa ccctgctgat 2700

gagcccactt caggggcccc agccccccag gaacttgggg ctgccaacca gcagggtgat 2760

ctttgcgagg tgagcctcgc cggcagtgtg gagcctgccc aaggagaggc cagggagggt 2820

tctccccgtg acaaccccac tgcacagcag atcatgcagc tgcttcgtga gatgcagaac 2880

ccccgggagc gcccaggctt gggcagcaac ccctgcattc ctttttttta ccgggctgac 2940

gagaatgatg aggtgaagat cactgtcatc 2970

<210> 57
<211> 461
<212> DNA
<213> Homo sapien

<400> 57
caggattgct ttgtccatct cctgctttca tttcaagtgc ataaacaaaa cctcaaaggg 60

cctgggaagg tgaggcaggc cagagtctgt gttctgtgtt gagtgtcaag ctatttgtta 120

ggaaggtctg caacaggcct tggtgtgggc tctgccagag actgttctga acacttgctt 180

gagatccgtg ccctgtaaaa tggatatgat gttttactga tgtctgtaat acatttgtaa 240

acttccaata aaatttgaat aaaagaaaaa taaaaaaaaa caacaaaaaa aagaaaaaag 300

aagcgcgggg cggtactgca ggggccatac gctggtgtcc cgtggggtgg acatgggtga 360

gatccgggtc aaaattccac ccaaactata gcgagcaatc ggagcatagc gacagagaag 420

agagagcgac acagagatgc agacgaccaa agaacaggaa g 461

<210> 58
<211> 1032
<212> DNA
<213> Homo sapien

<400> 58
cccataaaat atgactcact attgggagcc atactatttt ataagcttac ttcctgctga 60

caaaactagc tttcctcaag gaaatataaa ggaggggaaa gtcacatagt gttaggaaaa 120

cattcctgtg ttttgaatac gatgaatcca taggatagag aaaaatctgc ttgttctatt 180

ctgagagttc tctgagatat cccttcactc tgcttggcat ttggccattg atattcaaca 240

ggtcactgac caagcttttc taaatttttc agagagagtt acttaccaat aaggtctgtt 300

cttaaaccta cctagttgat tttcatatct ttccataaag tgtcatgatt ctatcataga 360

ccctgactta acattgtaag gactatgagt cctccatttt ttaattaatt ttttttagc 420

aaattaggac ttcggcaggt tttcctctcc taaactcatt ctttcctcca caggattgct 480

ttgtccatct cctgctttca tttcaagtgc ataaacaaaa cctcaaaggg cctgggaagg 540

tgaggcaggc cagagtctgt gttctgtgtt gagtgtcaag ctatttgtta agaaggtctg 600

caacaggcct ttggtgtggg ctctgccaga gactgttctg aacactttgc ttgagatccg 660

tgccctgtaa aatggatatg atgttttact gatgtctgta atacatttgt aaacttccaa 720

taaaatttga ataaaagaaa aaaaaaaaat caaaccacgg accacaagac acgagtacac 780

aaaaaccaag ggggcgcgcc cctcaagaat taccccccgag agagcgcaca aataagccac 840

cgccaccacc gtcattggac cggaggggcg ccacacaatg gacgccaatt aacacaagcc 900

gggccggcat taaaacacgc gcatcggaca ctgcgacacg agccgtggag gaaaccacac 960

gcggggcaca aaagcaagca caccggtaat ccccggacaa cacccagcta gtggtaccaa 1020

ccagcctcgg aa 1032

<210> 59
<211> 725
<212> DNA
<213> Homo sapien

<400> 59
gatgatcaac atatagggac atggttcatc tagatgcatg ctcgagcggc gcagtgtgat 60

ggatgtcgcg gccgaggtgt tggcacagaa gcccattgat ccctctggaa aatagggagt 120

ccctcctgag actggacagg ccgaacctgg ctctgtctcg taggcgccct gtgcatttcc 180

ttcccagcca gcgtcccagg cctggctcac agctgtggtg gcacatctga acttaagatc 240

ctggatttgg ttctgtcctg cccccaattt aaatagtcac aaatacagat gtagcagaag 300

aaaccccgca gcatccaagt cagttctgtg ggagtcgcat gttcctgtgt ctcacggcat 360

ggggcagagc cagtgagcat tcttgctgtc ctgccagtgt gtgggcctca gtgccacctg 420

ccattccctg gttttgattg cccaggcccc ctaacaccca caagggacag acttccacct 480

tcctttatcc attcacagtc cacgcctgcc ctgcagggac gctggtgggt gctggcctgg 540

gcttcctcac atacaactgc catgcttcct gcagggacgc tggtgggcgc tggcctggag 600

cttcctcacc taaaacccat gcttcctgca ggacgctggt gggtgctggc tggggcttct 660

cacatacaaa tgcatgcttc tgcaggacgc tggtatgcgg tgtctgacct acaaccatgc 720

tatcc 725

<210> 60
<211> 666
<212> DNA
<213> Homo sapien

<400> 60
cacaagggga aactcctcga ggctctggga gggacggagg gtttggtgac agagcgagag 60

ctaaaattga ggattcctga atccagatct tgcctcccat cagccatctt tctcccaata 120

aatttttgtt atgtgcaaaa aaaaaaaaaa aaaaaaacaa aaaaaaaaaa aaaaaaaaaa 180

aaaaagggtt ttttttttt ttttttttt ttttattttg tgggggggag agacggggag 240

ccaaaaagga gattttatta tacattttta gagaagagag agagagaaac aacaaggtag 300

aagcacaaac caagcaagcg acacaagaga gaaagggcgt gcctctcatc tacacaccac 360

actatcttct caaccacccc actcctcaca tcctattatc tcaacaaaca gggcgccgcc 420

gcagcgcaca caacaatagt cgaaagccgg gggggcgggg aaccactagg gcgggcgcaa 480

ccgcgggtgt agcagcgggg gcgggaaaaa agtggttact ccgcgggcac caaaatcctc 540

ccaacaacaa aattgaagca ggaacaaaaa gagtaacgac acaccaaacc accagcaaca 600

cagcacagcg acaacgaaca cacacagccg acacacacac cggcaccaag caacaaccat 660

cgcccg 666

<210> 61
<211> 1098
<212> DNA
<213> Homo sapien

<400> 61
aggagggtga ggacgtacaa ggagcatcgc aggcgaggaa acaacacaac ggccaggacc 60

taactgtggt gggaactgcc tttgtctcca cacactcgca atcaacatgc gtatttgcta 120

ttctcaaaca actcccttcc acccccttag gctgaaagga caaaggtggc ctttttctct 180

ccagccttga attgttccct gttggcttcc caagggccca tctgctggta cagtccacac 240

ttccaaagcc aagacccgag agggctttca ctgccccaag cctctctcct gtgaccttgg 300

gattctgtct tggcagaatc ctttgtcagc ggctcttgct ctgtccttcc tgtttggcca 360

cagctctttc aatcaatggg tattctagaa ccgcaggatg tcagagctgg aagggacgcg 420

ataccggttt acacaagggg aaactcctcg aggctctggg agggacggag ggttttggtg 480

acagagcgag agctaaaatt gaggattcct gaatccagat cttgcctccc atcagccatc 540

tttctcccaa taaatttttg ttttgtgcaa aaaaaaaaaa aaaaaaaaac aaaaaaaaaa 600

aaaaaaaaaa aaaaaaaggg ttttttttt ttttttttt ttttttattt tgtgggggg 660

agagacgggg agccaaaaag gagattttat tatacatttt tagagaagag agagagagaa 720

acaacaaggt agaagcacaa accaagcaag cgacacaaga gagaaagggc gtgcctctca 780

tctacacacc acactatctt ctcaaccacc ccactcctca catcctatta tctcaacaaa 840

cagggcgccg ccgcagcgca cacaacaata gtcgaaagcc gggggggcgg ggaaccacta 900

gggcgggcgc aaccgcgggt gtagcagcgg gggcgggaaa aaagtggtta ctccgcgggc 960

accaaaatcc tcccaacaac aaaattgaag caggaacaaa aagagtaacg acacaccaaa 1020

ccaccagcaa cacagcacag cgacaacgaa cacacacagc cgacacacac accggcacca 1080

agcaacaacc atcgcccg 1098

<210> 62
<211> 970
<212> DNA
<213> Homo sapien

<400> 62
gatatatagg cgaatgggct tctaatgcat gccgagcggc ggcaggtgat ggatgtgtta 60

taagaattat atccatatgt ctgccttggc tccaagtcat gcctcttaaa ataaaagata 120

caatccatac tagcatgaaa agtttccctc aacaggctat attaacatag tcatgagtgc 180

tgaccaaact caccgagctc agaggccagg catggcctga ggtgcagaat aggcctctgc 240

ctcccaagag ccctttcctt gccctgagca aggagtggtg ttccacaaac aaggctgctc 300

ttctaagcca acagtgtcag gcaggaagca gccataattt gccttgcatt ttcattccct 360

aatgtaaagg gatctgcatg gtcactctcc tgttctctga gccattgctc agggccagcc 420

aagatattat gagaacagat aatttacctt ggagccagag gccctccctg cctttagcaa 480

ggatgttcag ggacagacaa agagggcagt ggtggtgaat gttgttactg ccatgaggag 540

aaatggcagt aagaaatctt aactacaagc agccaatttc tcattccagg accctagcca 600

gaataataga cttctttttt ttttgagaca gagtttgctt ttgttgccct ggtggagtgc 660

agtggcgcaa tcttggtcac cgaactccac ttcccaggtt cagcaattct gctcagctcc 720

cgagtagctg ggattccggc atggcacagc ctggtatttg tatttagtag agaggggttc 780

tcatgtggtc aggcgttctc gaatccaggt ggtgatctcc gcccagttcc aaggtgggtt 840

cgggtggcca gctgttaggg atgattcttt gacttggtcc tccagtggtt tgtgcatgcc 900

tgatgagggg ggccctgaac gggggttttt gtgggccggg tggggcgggc cgggggcatg 960

tggttcgccc 970

<210> 63
<211> 1685
<212> DNA
<213> Homo sapien

<400> 63
catatgcacc actggatttt gcatacagcc tcatacagtg caaacaggat gtgacttgct 60

cagcttagtc atgtgattta tttaaaaaaa aaaaaaaaag aaacacaaaa cgataaatct 120

tctactcagg gtatagcaaa acaaaaaaat tccctttcca ccaaaaagcc tgaattgttc 180

caataagtta tctcatttgg aatgtttcat taatttgtgt tataggaaaa aaattgtgtg 240

tgtgtgttat aagaatttata tccatatgtc tgccttgggc tccaagtcat tgcctcttaa 300

aataaaagat acaatccata ctagcatgaa aggtttccct caacaggcta atattaacat 360

agtcatgagt gctgcccaaa ctcaccgagc tcagaggcca ggcatggcct gaggtgcaga 420

ataggcctct gcctcccaag agccctttcc ttgccctgag caaggagtgg tgttccacaa 480

acaaggctgc tcttctaagc caacagtgtc aggcaggaag cagccataat tttgccttgc 540

attttcattc cctaatgtaa agggatctgc attggtcact ctcctgttct ctgagccatt 600

gctcagggcc agccaagata ttattgagaa cagataattt accttggagc cagaggccct 660

ccctgccttt agcaaggatg ttcagggaca gacaaagagg gcagtggtgg tgaatgttgt 720

tactgccatg aggagaaatg gcagtaagaa atcttaacta caagcagcca atttctcatt 780

ccaggaccct agccagaata attgacttct ttttttttg agacagagtt tgcttttgtt 840

gccctggtgg agtgcagtgg cgcaatcttg gtcaccgaac tccacttccc aggttcagca 900

attctgctca gctcccgagt agctgggatt ccggcatggc acagcctggt atttgtattt 960

agtagagagg ggttctcatg tggtcaggcg ttctcgaatt cccactcagt gatctccccg 1020

cctcgccctc caagtgctgg gattacagcg tgaccaccgc gcctgccaac tgcttcagtt 1080

tcaagaaaga actagtcata acattccagg gcactcactg cctagttctc tcttgggatt 1140

taggggaaaa gacttcgaag tcaggtgatc taagaaatgc attccagttt ctctatggga 1200

tctcaactaa agctcgcatt attactctgg gcacagaaag tggtcactga gggccaaaca 1260

catttaaaag cttcatttcc ctaaaaagga aacctagact gctgacttct tacgtgaagc 1320

tgcctcagct gcactgataa ttctagaaca cttaaattcc aaaggaatga ctagggtgtt 1380

tatgaagtct acttggaacc cctgtcccac tttagaacac agggatcaac ggacttgacc 1440

atgttcattc aggggagaca ggtccttagg aaatcctgtc cagagtttta caacagagag 1500

gctaatgcag acacttttga agtgaggccc atgctatata ggaaaatgaa agttaggatt 1560

ttgagactct cagcctgttc tggaaaaatc ctggaagcaa gcggaatgaa atggtattat 1620

cttctctgac aagtggtcca gccacaggaa caggggggaac tgagcagaaa gcatatgtta 1680

tccag 1685

<210> 64
<211> 327
<212> DNA
<213> Homo sapien

<400> 64
ggtgatactc tatgccaatg tgcctctgat gctgctcgag cggcgccagt gtgatggata 60

cggagttagt ctgtttctaa atgaggggac agtatgtttc ttggggcctg aggacagctt 120

aataaagtag acaaacgaaa aaaaaaaaaa aaaaaaaaaa aaaaactttg ggctttatcc 180

ttggtccata gcttgtttac tctgtggtga tattgttccg tcaattccca cattaccagg 240

ggggacgctg cgcacggggg agagagggcg gggcggaagg cagcgaccgg agcgggcaag 300

cgcgggagga gagcacgacg gcgacac 327

<210> 65
<211> 5859
<212> DNA
<213> Homo sapien

<400> 65
gtgtcgccgt cgtgctctcc tcccgcgctt gcccgctccg gtcgctgcct tccgccccgc 60

cctctctccc ccgtgcgcag cgtccccccct ggtaatgtgg gaattgacgg aacaatatca 120

ccacagagta aacaagctat ggaccaagga taaagcccaa agttttttttt ttttttttt 180

ttttttttctt catttgtcta ctttattaag ctgtcctcag gccccaagaa acatactgtc 240

ccctcattta gaaacagact aactccgttt tcctccacta tcccctcccc tgtccttgat 300

ctgtagatcc tgttaagaca ggaaaaacag tgttggtcaa agggtacacg ctttcagtta 360

caagatgaac aagttctgaa tacgtaagat agaacatggg aggtgatgtg gccgggtgca 420

gtgactcacg cctgtaatcc cagcactttg ggaggccgag gtgggcggat catgaggtca 480

agggatcgag atcatcctgg ccaacatggt gaaaccccgt ctctactaaa aatacaaaaa 540

ttagctgggc atggtgggca cacgcctata gtcccagcta cttaggaggc tgaggcagga 600

gaattgcttg aacctgggag gcagaggttg cagtgagctg agatcgcgcc attgcactcc 660

agcctgggcg acaagagcaa aactccgtct cagaaaaaaa aaaaaaaaaa aaagagttga 720

tgtgttgaaa gacagagaag cgaagacaga gacgtggaaa gacagggaga gagacacgga 780

gagagacgca gaaggacaga gacgtggaga gagacgcaga gagacagaga cgtggagaga 840

cacagagaga cttggagaga gacaaagcaa gacaggacgg gagaacaagg acaagctcag 900

gtgcccctgg agccccagcc ctgccttcat gctcagcagg tgccctacct ggcccatcct 960

cccaaggtaa gcctcagccg gtgctgcagg cagtctgact cgcagtccct caagtgactt 1020

ccaaggagca tctgtagaaa agaagatggc ccaggtcctg cacgtgcctg ctcccttccc 1080

agggacccct ggcccagcct ccccacctgc cttccctgcc aaggaccccg atccacccta 1140

ctccgtggag acccccctatg gctaccgcct ggacctggac ttcctcaagt acgtggatga 1200

catcgagaag ggccacacgc tgcgacgcgt ggcagtgcag cgccgccccc gcctgagctc 1260

gctgccccgt ggccctggct cctggtggac gtccactgag tcgctgtgct ccaatgccag 1320

tggggacagc cgccactcag cctattccta ctgcggccgt ggcttctacc ctcagtatgg 1380

tgctctggag acccgcggtg gcttcaatcc gcgggtggag cgcacgctgc tggatgcccg 1440

tcgccgtctc gaggaccagg cggccacacc caccggcctg ggctccctga cccccagtgc 1500

ggccggctcg acagcctccc tggtgggcgt ggggttgcca cccccgacac cacggagttc 1560

aggactgtcc acaccggtgc ctcccagtgc cgggcacctg gcccacgtgc gggagcagat 1620

ggcgggtgcc ctgcggaagc tgcggcagct ggaggagcag gtgaagctga tccctgtgct 1680

ccaggtgaag ctctcggtgc tccaggagga aaagcggcag ctcacagtac aacttaagag 1740

ccagaagttc ctgggccacc ccacagcggg ccggggtcgc agcgagctct gcctggacct 1800

ccccgatccc ccagaggacc cagtggcact ggagacccgg agtgtgggca cctgggtccg 1860

agaacgggac ttgggcatgc ctgatgggga ggctgccctc gccgccaagg tcgctgtgct 1920

ggagacccag ctcaagaagg cgctgcagga gctgcaggca gctcaggccc ggcaggctga 1980

cccccagccc caggcctggc caccgccgga cagcccggtc cgcgtggata cagtccgggt 2040

ggtagaaggg ccacgggagg tggaggtggt ggccagcaca gccgctggcg cccccgcaca 2100

gcgggcccag agcctggagc cttacggcac agggctgagg gccctggcaa tgcctggtag 2160

gcctgagagc ccacctgtgt tccgcagcca ggaggtggtg gagacaatgt gcccagtgcc 2220

cgctgcagct accagcaacg tccatatggt gaagaagatt agcatcacag agcgaagctg 2280

cgatggagca gcaggcctcc cagaagttcc tgccgaatcg tcttcgtcac ccccggggtc 2340

cgaggtagcc tcccttacac agcctgagaa gagcacaggc cgagtgccca cccaggagcc 2400

cacccacagg gagcccacca ggcaagcagc ctcccaagag tccgaggagg ccgggggcac 2460

cggcgggccc ccggcaggcg tgcgatctat catgaaacgg aaagaggagg ttgcagaccc 2520

cacggcccac cggaggagcc tccagttcgt gggggtcaac ggcgggtatg agtcgtcatc 2580

cgaggactcc agcacagcag agaacatctc agacaacgac agcacagaga acgaggcccc 2640

agagccgagg gagagggttc cgagtgtggc cgaagccccc cagctcaggc ctgcagggac 2700

ggcagcggcc aagaccagcc ggcaggagtg tcagctgtct cgagaatctc agcacatacc 2760

cactgctgag ggggcatcag gatcaaacac ggaggaggag atcaggatgg agctaagccc 2820

tgacctcatc tcagcctgct tggccctgga aaagtacctg gacaatccca acgccctcac 2880

agagcgggag ctgaaagtgg cctacaccac agtgctgcag gagtggctgc gcctggcctg 2940

ccgcagcgac gcacaccccg agctggtgcg gcggcacctg gtcacgttcc gggccatgtc 3000

tgcgcggctg ctggactacg tggtcaacat cgccgacagc aacggcaaca cagccctgca 3060

ctactccgtg tctcatgcca acttccccgt ggtgcagcag ctgctcgaca gcggtgtctg 3120

caaggtggac aaacagaacc gtgctggcta cagccctatt atgctcaccg ccctggccac 3180

cctgaagacc caggacgaca tcgagactgt ccttcagctc ttccggcttg gcaacatcaa 3240

tgccaaagcc agccaggcag gacagacggc cctgatgctg gccgtcagcc acgggcgggt 3300

ggacgttgtc aaagccctgc tggcctgtga ggcagatgtc aacgtgcaag atgatgacgg 3360

ctccacggcc ctcatgtgcg cctgtgagca cggccacaag gagatcgcgg ggctgctgct 3420

ggccgtgccc agctgtgaca tctcactcac agatcgcgat gggagcacag ctctgatggt 3480

ggccttggac gcagggcaga gtgagattgc gtccatgctg tattcccgca tgaacatcaa 3540

gtgctcgttt gccccaatgt cagatgacga gagccctaca tcatcctcgg cagaagagta 3600

gccgtgaggg aggcggggac cagccagacc gggagcaaac cgtcccttgt ccccgtctcc 3660

tccctgttcc cgttcctccc tggcccaccc cactcacact ccccaaggcc cacggctcaa 3720

aggcaagcga gctctccctc tgcttccctg ggggagcccc aacggccaca ggactccagc 3780

tccaagtggg ttttcttggc tcccctgttc aaagtggcca cagcgcagac cgaagcaaaa 3840

ttcttgtata cattggcgcc agggctgatg ctggggtgtg ggttttatga agaacattga 3900

gaacaatcag ctggtaatta tggatggagg aagagggaga ggaaaaaaat attgtatttt 3960

tgaatcattg ttgcaggagg gggtgggaat cttaggattt gttgccagat ttgaaagtca 4020

ctggaacttg catattttca ttttaatcct aagtgttatt acgcaccagt tggggttcac 4080

ccttcatccc tcacatttaa ttgtctgata tagaatagtg ttgtgtccac tgccccgcta 4140

gacggctttc ttaggggaat tttcttctgg ttgtttcaca agacagattc tgtccttgtc 4200

acccgggaca gaaaactcag tcttttcacc ctcattcaga tgaagggact caggacaggc 4260

tctgtgactt acagggaccc aatcaattca caatgagaaa ttaccggcca ggcgtggtga 4320

ctcacgtctg taatcccagc actttgggag ggcaaggcaa gagcttgagc ttgagcctag 4380

acgttaaaga ccagcctggg caacacagca agacccatct ctacaagaaa tttaaaaact 4440

agccaggcgt ggtggtgcgc gcctgtagtc ccagctactt gggaggctga gccctggagg 4500

tcgaggctac agtgagctat gatcacacca ttgcacttca gcctgggcga cacagcgaga 4560

ccctgtctca agaaagaaaa aaaaaagaga caaattaccc agaaacccct cccttcccca 4620

catggaggcc ttggcaaatg ttaattttcc tagaaaatcc ttcagacctg aagacgcagg 4680

aaaagaatct ggctctcagg gtggcttctg cgtccccgcc gccaggcccc agactatggt 4740

cacagggccg tcctgttcct ccccgggact ccagaatttc tctcctcaaa ggaaagaaaa 4800

cagggcatgc gcttgttggc aaaacgcagg gccggctccc aaaaacccca tgtgtgtacg 4860

attaaaagtt ggccgtcccc aggcctccca gcgcaaactt aaagagacag ggctttgctg 4920

aaaaccaaac atgggccagc tgggcttttt aacaacctag agactttccg gagctgcctg 4980

gaacagagcc tgtgggaaac ggggcttgcc agagacactc acagtttcct tcatggcctg 5040

ttttggtccc ctaagaatct ccacatcatt gtctttcttg tgccttttcc ttggtgagca 5100

acagaaaggg aagggttcca agcctctaaa aatgtgcttt gtgatcagga gtgcgctcca 5160

aaccaaatac gcgcgctgcc ctttcgaggc cagtgagctc agcctccaag gctttaaagc 5220

cacatttcag caagagaaag cgctgagagc tcgcaggttc attaaagaag gcaaagcact 5280

ggtttctctc cttagaaaag taggtttctt ggcttgatgt agactggctt gctttgattt 5340

ttagtgaagg gaatgtacgt aaaacaaaat agggcttggc tggtcaaagg agacaagcag 5400

gatggatgga tggatggatg aatagataga tggtgtttgc atgtaaattg cagagaaaac 5460

aaaaccaaag ctgattggaa acaattaatt gtgggtgtct gaggggggaag gtcgcagctt 5520

tgggcagctt tgagaagcgg tacaagagct ctgtgcctgt gtgtccagcc ctggagccag 5580

ccagtgcatt tattttaagc tcttagaagc aactccttgg cccaggaatg cgtgacccct 5640

gagatgggtc cacgcatctc tctacacgtc cttctctccg tgggatactg gactcgtgcc 5700

tctgcgccca ttctcttctc acgcatatcc atgagcttta atttcacttt ctgatcacgg 5760

tacgtccata aagccagtat tacacttaaa tgaagtattc ttttttgtaa tcgttttttt 5820

tagaaggtaa acaaatttaa taaagctacc aataatgtt 5859

<210> 66
<211> 93
<212> PRT
<213> Homo sapien

<400> 66

Met Gly Gly Asn Val Gly Arg Glu Thr Asn Val Pro Pro Gly Ala Ser
1 5 10 15

Phe Gly Pro Trp Val Pro Pro Ala Phe Phe Phe Phe Cys Phe Phe Val
20 25 30

Phe Phe Phe Lys Arg Arg Ile Leu Gly Phe Phe Gly Glu Thr Lys Ala
35 40 45

Asp Ile Lys Ser Tyr Lys Asp Phe Arg Phe Ser Phe Thr Lys Lys Val
50 55 60

Ile His Ile Leu His Tyr Thr Arg Tyr Asp Ile Asn Thr Gly Lys Tyr
65 70 75 80

Tyr Val His Cys Lys Glu Lys Gly Lys Ile Glu Thr Tyr
85 90

<210> 67
<211> 59
<212> PRT
<213> Homo sapien

<400> 67

Met Gly Lys Lys Ala His Arg His Leu Gln Phe Thr Ser Phe Lys Phe
1 5 10 15

Leu Lys Lys Thr Pro Gln Lys Lys Pro Phe Leu Pro Gly Lys Ala His
20 25 30

Glu Ile Asn Tyr Arg Ile Glu Leu Tyr Asn Ser Thr Ser Thr Ser Leu
35 40 45

Thr Leu Met Cys Phe Ala Lys Asn Leu Glu Lys
50 55

<210> 68
<211> 59
<212> PRT
<213> Homo sapien

<400> 68

Met Ser Ile Tyr Ser Phe Ile Leu Val Lys Asn Ile Arg Gln Ser Arg
1 5 10 15

Gly Arg Phe Lys Ser Glu Lys Lys Lys Lys Lys Lys Lys Lys Ser Ala
20 25 30

Gly Gly Thr Ser Gly Pro Lys Gly Ser Arg Gly Glu Leu Val Ser Arg
35 40 45

Pro Lys Phe Pro Pro Asn Phe Pro Pro Lys Gly
50 55

<210> 69
<211> 55
<212> PRT
<213> Homo sapien

<400> 69

Met Thr Ile Leu Asn Tyr Ser Ile Asn Met Arg Cys Trp Leu Lys Ser
1 5 10 15

Phe Ser Arg Leu Leu Met Ser Thr Ser Val Leu Val Phe Leu Gly Thr
20 25 30

Ser Tyr Phe Tyr Leu Gly Phe Trp Pro Tyr Leu Ser Ser Ile Thr Ser
35 40 45

Pro Glu Thr Ser His Gly Asn
50 55

<210> 70
<211> 69
<212> PRT
<213> Homo sapien

<400> 70

Met Ser Val Phe Phe Cys Val Lys Thr Pro Asp Thr Lys Thr Thr His

1 5 10 15

Lys Thr Asn Lys Arg Lys Glu Asn Val Ala Arg Ile Leu Val Ser Leu
20 25 30

Thr Val Glu Asp Pro Asp Gln Ala Val Gln Asn Val Ala His Gly Thr
35 40 45

Glu Arg Thr Gly Val Thr Thr Glu Ile Lys Phe Val Gly Leu Gly Val
50 55 60

Val Ala Pro Ser Gly
65

<210> 71
<211> 59
<212> PRT
<213> Homo sapien

<400> 71

Met Leu Ala Asp Ile Gly Val Leu Ile His Met Lys Trp Ile Asp Thr
1 5 10 15

Ser Ser Arg His His Thr Ala Val Gln Ser Ile Gln Gly Arg Glu Ala
20 25 30

Thr Ser Arg Leu Thr Thr Phe Leu Ala Gly Ser Gly Glu Leu Cys Pro
35 40 45

Arg Lys Pro Thr Arg Arg Ser Gly Thr Glu Glu
50 55

<210> 72
<211> 50
<212> PRT
<213> Homo sapien

<400> 72

Met Phe Cys Ser Glu Asn Thr Leu Pro Gln Asp Ile Leu Gln Leu Ser
1 5 10 15

Tyr Cys Ile Gln Leu Ser Ala Gln Val Leu Thr Asp Glu Thr Cys His
20 25 30

Pro Tyr Ser Thr Pro Cys Ser Ala Leu Leu Asn Ser Asn Ala His Met
35 40 45

Ala Pro
50

<210> 73
<211> 74
<212> PRT
<213> Homo sapien

<400> 73

Met Lys Gln Arg Ile Ser Lys Glu Thr Thr Lys Asp Ile Gly Asn Ser
1 5 10 15

Gln Lys Pro His Ala Asp Ala Glu Leu Gly Val Lys Asp Cys His Thr
20 25 30

Val Ser Asn Cys Arg Gly Val Cys His Ile Asp Ala Phe His Thr Leu
35 40 45

Glu Val Ala Arg Ala Ser Trp Val Thr Leu Pro Gln Arg Lys Asp Arg
50 55 60

Cys Val Pro Gly Gln Cys Arg Gly Glu Met
65 70

<210> 74
<211> 133
<212> PRT
<213> Homo sapien

<400> 74

Met Lys Ser Gln Glu Arg Met Asn Ser Cys Asp Gln Leu Gln Lys Thr
1 5 10 15

Gln Ala Asp Ser Ile Leu Arg Asp Thr Leu Tyr His Phe Gly Arg Ser
20 25 30

Pro Thr His Leu Gly Lys Thr Gly Met Ser Leu Arg Gly Ser Gly Arg
35 40 45

Ser Ser Arg Trp Leu Thr Val Val Gly Ala Ala Val Val Ala Val Val
50 55 60

Ala Ala Asp Ser Gly Phe Ser Ile Arg Gly Phe Ile Ile Ser Arg Thr
65 70 75 80

```
Ser Ser Trp Ile Arg Val Ser Trp Ile Ser Cys Tyr Ser Asp Leu Trp
                85                  90                  95

Ala Glu Thr Thr Asn Asp Gly Thr Pro Gln Ser Thr Ser Pro Thr Ser
            100                 105                 110

Ala Ile His Thr Leu Ala Pro Arg Arg His Asp Leu Glu Ala His Arg
        115                 120                 125

Leu Ser Gly Tyr His
    130
```

<210> 75
<211> 72
<212> PRT
<213> Homo sapien

<400> 75

```
Met Trp Ser Val Ser Pro Cys Ser Leu Pro Glu Gln Cys Leu Arg Phe
1               5                   10                  15

Glu Trp Asp Pro Thr Phe Val Asn Glu Ile Tyr His Leu Pro Arg Gln
            20                  25                  30

Asn Asn Arg Phe Cys Pro Arg Cys Cys Asp Val Thr Met Val Ala Ile
        35                  40                  45

Thr Ala Ile Thr Tyr Asn Tyr Trp His Thr Tyr Asp Glu Ser Arg Thr
    50                  55                  60

Gly Pro Lys Cys Phe Leu Thr Met
65                  70
```

<210> 76
<211> 93
<212> PRT
<213> Homo sapien

<400> 76

```
Met Ser Leu Cys Cys Asp Gly Pro Phe Pro Ser Leu Phe Gly Tyr Pro
1               5                   10                  15

Pro Leu Thr Ile Leu Ile His Val Leu Phe Gln Lys Val Ser Pro Ile
            20                  25                  30
```

Lys Trp His Leu Gly Thr Thr Met Ala Gly Ile Ala Leu Ala Met Asn
35 40 45

Ser Thr Val Val Thr Leu Ser His Ser Arg Ala Val His Phe Ile Met
50 55 60

Asn Asp Leu Arg Ile Ser Pro Gly Lys Ser Pro Arg Gln Ala Leu Pro
65 70 75 80

Leu Leu Leu Ala Leu Gln Cys Glu Val Ser Trp Glu Arg
85 90

<210> 77
<211> 500
<212> PRT
<213> Homo sapien

<400> 77

Met Lys Cys Thr Ala Arg Glu Trp Leu Arg Val Thr Thr Val Leu Phe
1 5 10 15

Met Ala Arg Ala Ile Pro Ala Met Val Val Pro Asn Ala Thr Leu Leu
20 25 30

Glu Lys Leu Leu Glu Lys Tyr Met Asp Glu Asp Gly Glu Trp Trp Ile
35 40 45

Ala Lys Gln Arg Gly Lys Arg Ala Ile Thr Asp Asn Asp Met Gln Ser
50 55 60

Ile Leu Asp Leu His Asn Lys Leu Arg Ser Gln Val Tyr Pro Thr Ala
65 70 75 80

Ser Asn Met Glu Tyr Met Thr Trp Asp Val Glu Leu Glu Arg Ser Ala
85 90 95

Glu Ser Trp Ala Glu Ser Cys Leu Trp Glu His Gly Pro Ala Ser Leu
100 105 110

Leu Pro Ser Ile Gly Gln Asn Leu Gly Ala His Trp Gly Arg Tyr Arg
115 120 125

Pro Pro Thr Phe His Val Gln Ser Trp Tyr Asp Glu Val Lys Asp Phe
130 135 140

Ser Tyr Pro Tyr Glu His Glu Cys Asn Pro Tyr Cys Pro Phe Arg Cys
145 150 155 160

Ser Gly Pro Val Cys Thr His Tyr Thr Gln Val Val Trp Ala Thr Ser
165 170 175

Asn Arg Ile Gly Cys Ala Ile Asn Leu Cys His Asn Met Asn Ile Trp
180 185 190

Gly Gln Ile Trp Pro Lys Ala Val Tyr Leu Val Cys Asn Tyr Ser Pro
195 200 205

Lys Gly Asn Trp Trp Gly His Ala Pro Tyr Lys His Gly Arg Pro Cys
210 215 220

Ser Ala Cys Pro Pro Ser Phe Gly Gly Gly Cys Arg Glu Asn Leu Cys
225 230 235 240

Tyr Lys Glu Gly Ser Asp Arg Tyr Tyr Pro Pro Arg Glu Glu Glu Thr
245 250 255

Asn Glu Ile Glu Arg Gln Gln Ser Gln Val His Asp Thr His Val Arg
260 265 270

Thr Arg Ser Asp Asp Ser Ser Arg Asn Glu Val Ile Ser Ala Gln Gln
275 280 285

Met Ser Gln Ile Val Ser Cys Glu Val Arg Leu Arg Asp Gln Cys Lys
290 295 300

Gly Thr Thr Cys Asn Arg Tyr Glu Cys Pro Ala Gly Cys Leu Asp Ser
305 310 315 320

Lys Ala Lys Val Ile Gly Ser Val His Tyr Glu Met Gln Ser Ser Ile
325 330 335

Cys Arg Ala Ala Ile His Tyr Gly Ile Ile Asp Asn Asp Gly Gly Trp
340 345 350

Val Asp Ile Thr Arg Gln Gly Arg Lys His Tyr Phe Ile Lys Ser Asn
355 360 365

Arg Asn Gly Ile Gln Thr Ile Gly Lys Tyr Gln Ser Ala Asn Ser Phe
370 375 380

Thr Val Ser Lys Val Thr Val Gln Ala Val Thr Cys Glu Thr Thr Val
385 390 395 400

Glu Gln Leu Cys Pro Phe His Lys Pro Ala Ser His Cys Pro Arg Val
405 410 415

Tyr Cys Pro Arg Asn Cys Met Gln Ala Asn Pro His Tyr Ala Arg Val
420 425 430

Ile Gly Thr Arg Val Tyr Ser Asp Leu Ser Ser Ile Cys Arg Ala Ala
435 440 445

Val His Ala Gly Val Val Arg Asn His Gly Gly Tyr Val Asp Val Met
450 455 460

Pro Val Asp Lys Arg Lys Thr Tyr Ile Ala Ser Phe Gln Asn Gly Ile
465 470 475 480

Phe Ser Glu Ser Leu Gln Asn Pro Pro Gly Gly Lys Ala Phe Arg Val
485 490 495

Phe Ala Val Val
500

<210> 78
<211> 51
<212> PRT
<213> Homo sapien

<400> 78

Met Val Thr Thr Gln Asn Leu Arg Leu Thr Ile Val Glu Val Arg Gly
1 5 10 15

Gln Gly Ala Gly Arg Ala Gly Ser Phe Leu Ser Ser Ile Met Gly Ala
20 25 30

Ala Gly Arg Ile Gln Phe Leu Ala Gly Leu Gly Arg Arg Ser Pro Val
35 40 45

Pro Ala Ala
50

<210> 79
<211> 50
<212> PRT

<213> Homo sapien

<400> 79

Met Val Phe Tyr Tyr Tyr Tyr Gly Phe Lys Lys Ser Asn Phe Ile
1 5 10 15

Ser Phe Cys Lys Glu Leu Ser Asn Ile Leu Tyr Arg Phe Cys Glu Arg
20 25 30

Thr Tyr Phe Leu Thr Val Ile Phe Ile Ser Phe Lys Ile Phe Val Ser
35 40 45

His Leu
50

<210> 80
<211> 229
<212> PRT
<213> Homo sapien

<400> 80

Met Ala Glu Glu Met Glu Ser Ser Leu Glu Ala Ser Phe Ser Ser Ser
1 5 10 15

Gly Ala Val Ser Gly Ala Ser Gly Phe Leu Pro Pro Ala Arg Ser Arg
20 25 30

Ile Phe Lys Ile Ile Val Ile Gly Asp Ser Asn Val Gly Lys Thr Cys
35 40 45

Leu Thr Tyr Arg Phe Cys Ala Gly Arg Phe Pro Asp Arg Thr Glu Ala
50 55 60

Thr Ile Gly Val Asp Phe Arg Glu Arg Ala Val Glu Ile Asp Gly Glu
65 70 75 80

Arg Ile Lys Ile Gln Leu Trp Asp Thr Ala Gly Gln Glu Arg Phe Arg
85 90 95

Lys Ser Met Val Gln His Tyr Tyr Arg Asn Val His Ala Val Val Phe
100 105 110

Val Tyr Asp Met Thr Asn Met Ala Ser Phe His Ser Leu Pro Ser Trp
115 120 125

Ile Glu Glu Cys Lys Gln His Leu Leu Ala Asn Asp Ile Pro Arg Ile
130 135 140

Leu Val Gly Asn Lys Cys Asp Leu Arg Ser Ala Ile Gln Val Pro Thr
145 150 155 160

Asp Leu Ala Gln Lys Phe Ala Asp Thr His Ser Met Pro Leu Phe Glu
165 170 175

Thr Ser Ala Lys Asn Pro Asn Asp Asn Asp His Val Glu Ala Ile Phe
180 185 190

Met Thr Leu Ala His Lys Leu Lys Ser His Lys Pro Leu Met Leu Ser
195 200 205

Gln Pro Pro Asp Asn Gly Ile Ile Leu Lys Pro Glu Pro Lys Pro Ala
210 215 220

Met Thr Cys Trp Cys
225

<210> 81
<211> 42
<212> PRT
<213> Homo sapien

<400> 81

Met Asn Val Phe Lys Ile Tyr Asn Arg Thr Gln Ser Gly Arg Val Phe
1 5 10 15

Phe Gly Gly Arg Gly Leu Phe Ser Asn Ser Arg Trp His Ile Ser Gly
20 25 30

Gln Gln Tyr Phe Leu Thr His Ser Asn Gln
35 40

<210> 82
<211> 56
<212> PRT
<213> Homo sapien

<400> 82

Met Tyr Leu Lys Glu Lys Tyr Pro Asp Leu Lys Pro Thr Ala Asp Val
1 5 10 15

Ala Asn Phe His Thr Thr Ala Gly His Gly Ser Leu Leu Thr Thr His

20 25 30

Cys His Leu Arg Leu Cys Leu Cys Phe Ile Gln Arg Glu Arg Gly Gly
35 40 45

Leu Lys Gly Met Leu Pro Gly Gly
50 55

<210> 83
<211> 72
<212> PRT
<213> Homo sapien

<400> 83

Met Leu Ser Pro Phe Leu Leu Ile Asn Asn Leu Tyr Tyr Lys Lys Lys
1 5 10 15

Lys Lys Lys Lys Lys Arg Arg Gly Gly Asn Gln Gly Pro Ile Arg Gly
20 25 30

Phe Pro Gly Gly Glu Trp Val Thr Arg Ser Gln Phe His Thr Phe Ala
35 40 45

Arg Gln Gln Thr Gly Glu Glu Ala Gly Pro Arg Arg Glu Ala Arg Gln
50 55 60

Glu Gln Ala His Arg Glu Thr Glu
65 70

<210> 84
<211> 27
<212> PRT
<213> Homo sapien

<400> 84

Met His Val Glu Arg Arg Ser Val Met Asp Ala Trp Ser Arg Arg Gly
1 5 10 15

Ala Gly Lys Tyr Thr Asp Ile Lys Asp Gln Ile
20 25

<210> 85
<211> 292
<212> PRT
<213> Homo sapien

<400> 85

```
Met Asn Arg Phe Gly Thr Arg Leu Val Gly Ala Thr Ala Thr Ser Ser
1               5                  10                  15

Pro Pro Pro Lys Ala Arg Ser Asn Glu Asn Leu Asp Lys Ile Asp Met
            20                  25                  30

Ser Leu Asp Asp Ile Ile Lys Leu Asn Arg Lys Glu Gly Lys Lys Gln
        35                  40                  45

Asn Phe Pro Arg Leu Asn Arg Arg Leu Leu Gln Gln Ser Gly Ala Gln
    50                  55                  60

Gln Phe Arg Met Arg Val Arg Trp Gly Ile Gln Gln Asn Ser Gly Phe
65                  70                  75                  80

Gly Lys Thr Ser Leu Asn Arg Arg Gly Arg Val Met Pro Gly Lys Arg
                85                  90                  95

Arg Pro Asn Gly Val Ile Thr Gly Leu Ala Ala Arg Lys Thr Thr Gly
            100                 105                 110

Ile Arg Lys Gly Ile Ser Pro Met Asn Arg Pro Pro Leu Ser Asp Lys
        115                 120                 125

Asn Ile Glu Gln Tyr Phe Pro Val Leu Lys Arg Lys Ala Asn Leu Leu
    130                 135                 140

Arg Gln Asn Glu Gly Gln Arg Lys Pro Val Ala Val Leu Lys Arg Pro
145                 150                 155                 160

Ser Gln Leu Ser Arg Lys Asn Asn Ile Pro Ala Asn Phe Thr Arg Ser
                165                 170                 175

Gly Asn Lys Leu Asn His Gln Lys Asp Thr Arg Gln Ala Thr Phe Leu
            180                 185                 190

Phe Arg Arg Gly Leu Lys Val Gln Ala Gln Leu Asn Thr Glu Gln Leu
        195                 200                 205

Leu Asp Asp Val Val Ala Lys Arg Thr Arg Gln Trp Arg Thr Ser Thr
    210                 215                 220

Thr Asn Gly Gly Ile Leu Thr Val Ser Ile Asp Asn Pro Gly Ala Val
225                 230                 235                 240
```

Gln Cys Pro Val Thr Gln Lys Pro Arg Leu Thr Arg Thr Ala Val Pro
245 250 255

Ser Phe Leu Thr Lys Arg Glu Gln Ser Asp Val Lys Lys Val Pro Lys
260 265 270

Gly Val Pro Leu Gln Phe Asp Ile Asn Ser Val Gly Lys Gln Thr Arg
275 280 285

Ile Thr Leu Lys
290

<210> 86
<211> 34
<212> PRT
<213> Homo sapien

<400> 86

Met Val Phe Lys Glu Leu Ser Val Leu Pro Arg Cys Phe Trp Gly Ser
1 5 10 15

Pro Val Phe His Ser Val Ile Pro Phe Lys Arg Leu Ser Lys Ser Leu
20 25 30

Phe Asn

<210> 87
<211> 26
<212> PRT
<213> Homo sapien

<400> 87

Met His Thr Phe Thr Gly Lys His Asn Ser Phe Ser Leu Arg Lys Asn
1 5 10 15

Ala Glu Tyr Leu Leu Gln Leu Arg Lys Ile
20 25

<210> 88
<211> 129
<212> PRT
<213> Homo sapien

<400> 88

His Met Phe Glu Asp Phe Ser Phe Pro Phe Ala Ile Phe Leu Phe Phe
1 5 10 15

Leu Arg Arg Arg Ser Ala Leu Thr Pro Arg Leu Glu Ala Ser Gly Ala
20 25 30

Ile Leu Ala Tyr Cys Asn Leu His Pro Pro Gly Ser Ser Asp Ser Pro
35 40 45

Ala Ser Ala Ser Gly Val Ala Gly Ile Thr Gly Ala Arg His His Val
50 55 60

Arg Leu Ile Phe Val Phe Ser Val Glu Thr Gly Phe Cys Tyr Val Gly
65 70 75 80

Gln Ala Gly Leu Lys Leu Leu Thr Ser Ser Asp Pro Pro Ala Ser Ala
85 90 95

Ser Gln Ser Val Arg Ile Thr Gly Val Ser His Arg Ala Arg Leu Lys
100 105 110

Ile Phe Leu Asn Cys Asn Lys Tyr Ser Ala Phe Phe Glu Ser Leu Tyr
115 120 125

Leu

<210> 89
<211> 15
<212> PRT
<213> Homo sapien

<400> 89

Met Ala Thr Leu Ala Gly Tyr Phe Leu Ala Lys Phe Leu Leu Arg
1 5 10 15

<210> 90
<211> 71
<212> PRT
<213> Homo sapien

<400> 90

Met Lys His Gly Ser Phe Tyr Phe Thr Val Ser Asn Leu Ile Ala Ser
1 5 10 15

His Leu Lys Ser Ala Lys Ile Glu Leu Pro Lys Lys Cys Tyr Met Pro

20 25 30

Lys Gly Ala His Asn Tyr Leu Met Ala Lys Leu Ile Lys Leu Thr Ser
35 40 45

Pro Lys Ser Asp Ser Arg Asp Leu Leu Cys Pro Ser Leu Trp Cys Phe
50 55 60

Phe Ala Leu His Ile Cys Phe
65 70

<210> 91
<211> 35
<212> PRT
<213> Homo sapien

<400> 91

Met Leu Ala Arg Leu Leu Leu Met Ile Lys Ser Leu Asp Pro His Thr
1 5 10 15

Arg Phe Ala Met Val Thr Leu Ser Arg Thr Glu Ile Pro Leu Val Leu
20 25 30

Tyr Lys Arg
35

<210> 92
<211> 48
<212> PRT
<213> Homo sapien

<400> 92

Met Phe Thr Ser Thr Thr Leu Asn Gln Leu Leu Ser Ile Leu Tyr Ile
1 5 10 15

Phe Tyr Ser Ile Phe Phe Ser Asn Phe Leu His Phe Pro Met Ser Leu
20 25 30

Lys Phe Ser Val Asn Val Asn Phe Lys Asn Cys Thr Val Trp Leu Phe
35 40 45

<210> 93
<211> 67
<212> PRT
<213> Homo sapien

<400> 93

Met Cys Met Ser Arg Phe Glu Ser Leu Gly Cys Arg Phe Val Leu Pro
1 5 10 15

Trp Gln Arg Lys Arg Ser Leu Trp Gly Gly Glu Leu Phe Leu Val Ile
20 25 30

Ser Gly Lys Arg His Ile Glu Thr Leu Tyr Glu Trp Gly Phe Cys Phe
35 40 45

Lys Cys Trp Lys Ile Arg Ala Gly Ile Thr Cys Leu Gln Val Val Pro
50 55 60

Ser Leu Val
65

<210> 94
<211> 145
<212> PRT
<213> Homo sapien

<400> 94

Met Leu Pro Ala Gly Thr Leu Val Gly Ala Gly Leu Gly Val Pro His
1 5 10 15

Pro Gln Thr Pro Cys Phe Leu Gln Gly His Trp Trp Val Leu Ala Trp
20 25 30

Gly Phe Leu Thr His Lys His His Ala Ser Cys Arg Asp Val Asp Gly
35 40 45

Arg Trp Pro Gly Arg Ser Ser His Thr Thr Ala Met Leu Pro Ala Gly
50 55 60

Thr Leu Val Gly Ala Gly Leu Gly Leu Pro His Ile Gln Thr Pro Cys
65 70 75 80

Phe Leu Gln Gly Arg Trp Cys Ala Leu Ala Trp Gly Phe Leu Thr Tyr
85 90 95

Lys Pro His Ala Ser Tyr Arg Ala Arg Trp Trp Thr Ala Gly Pro Glu
100 105 110

Ala Ser Ser His Thr Ile Ala Ile Leu Pro His Gly Thr Leu Ala Ala
115 120 125

Arg Thr Gly Leu Gly Leu Pro His Pro Gln Thr Pro Cys Leu Pro Ile
130 135 140

Asp
145

<210> 95
<211> 48
<212> PRT
<213> Homo sapien

<400> 95

Met Gly Val Tyr Ser Gly Ala Gln Asn Ile Pro Thr His Asn Thr Ile
1 5 10 15

Ser Ser Gly Thr Ala Lys Lys Gly Glu Asn Arg Lys Gln Glu Asn Arg
20 25 30

Lys Lys Lys Arg Lys Lys Lys Lys Asn Arg Lys Lys Lys Lys Asn Glu
35 40 45

<210> 96
<211> 71
<212> PRT
<213> Homo sapien

<400> 96

Met Ala Gly Gly Ala Lys Glu Leu Pro Arg Ala Ser Phe Ile Arg Ala
1 5 10 15

Leu Ile Leu Cys Lys Arg Ala Glu Ser Ser Gly Pro Asn Arg Phe Pro
20 25 30

Lys Leu Leu Thr Leu Gly Met Arg Val Gln Tyr Thr Asn Phe Trp Gly
35 40 45

Thr Gln Thr Phe Arg Pro Gln Gln Tyr Pro Asn Tyr Ile Arg Asp Leu
50 55 60

Lys Ser Thr Thr Lys Asn Lys
65 70

<210> 97
<211> 291
<212> PRT
<213> Homo sapien

<400> 97

```
Met Leu Arg Arg Glu Ala Arg Leu Arg Arg Glu Tyr Leu Tyr Arg Lys
1               5                  10                  15

Ala Arg Glu Glu Ala Gln Arg Ser Ala Gln Glu Arg Lys Glu Arg Leu
            20                  25                  30

Arg Arg Ala Leu Glu Glu Asn Arg Leu Ile Pro Thr Glu Leu Arg Arg
        35                  40                  45

Glu Ala Leu Ala Leu Gln Gly Ser Leu Glu Phe Asp Asp Ala Gly Gly
    50                  55                  60

Glu Gly Val Thr Ser His Val Asp Asp Glu Tyr Arg Trp Ala Gly Val
65                  70                  75                  80

Glu Asp Pro Lys Val Met Ile Thr Thr Ser Arg Asp Pro Ser Ser Arg
                85                  90                  95

Leu Lys Met Phe Ala Lys Glu Leu Lys Leu Val Phe Pro Gly Ala Gln
            100                 105                 110

Arg Met Asn Arg Gly Arg His Glu Val Gly Ala Leu Val Arg Ala Cys
        115                 120                 125

Lys Ala Asn Gly Val Thr Asp Leu Leu Val Val His Glu His Arg Gly
    130                 135                 140

Thr Pro Val Gly Leu Ile Val Ser His Leu Pro Phe Gly Pro Thr Ala
145                 150                 155                 160

Tyr Phe Thr Leu Cys Asn Val Val Met Arg His Asp Ile Pro Asp Leu
                165                 170                 175

Gly Thr Met Ser Glu Ala Lys Pro His Leu Ile Thr His Gly Phe Ser
            180                 185                 190

Ser Arg Leu Gly Lys Arg Val Ser Asp Ile Leu Arg Tyr Leu Phe Pro
        195                 200                 205

Val Pro Lys Asp Asp Ser His Arg Val Ile Thr Phe Ala Asn Gln Asp
    210                 215                 220
```

Asp Tyr Ile Ser Phe Arg His His Val Tyr Lys Lys Thr Asp His Arg
225 230 235 240

Asn Val Glu Leu Thr Glu Val Gly Pro Arg Phe Glu Leu Lys Leu Tyr
245 250 255

Met Ile Arg Leu Gly Thr Leu Glu Gln Glu Ala Thr Ala Asp Val Glu
260 265 270

Trp Arg Trp His Pro Tyr Thr Asn Thr Ala Arg Lys Arg Val Phe Leu
275 280 285

Ser Thr Glu
290

<210> 98
<211> 39
<212> PRT
<213> Homo sapien

<400> 98

Met Ser Ile Arg Ala Trp Phe Pro Leu Ser Cys Arg Ala Ala His Val
1 5 10 15

Met Asp Pro Gly Arg Tyr Trp Thr Pro Gly Met Leu Thr Ala Thr Cys
20 25 30

Arg Gln Glu Thr Ser Val Gln
35

<210> 99
<211> 174
<212> PRT
<213> Homo sapien

<400> 99

Met Ser Phe Lys Arg Glu Gly Asp Asp Trp Ser Gln Leu Asn Val Leu
1 5 10 15

Lys Lys Arg Arg Val Gly Asp Leu Leu Ala Ser Tyr Ile Pro Glu Asp
20 25 30

Glu Ala Leu Met Leu Arg Asp Gly Arg Phe Ala Cys Ala Ile Cys Pro
35 40 45

His Arg Pro Val Leu Asp Thr Leu Ala Met Leu Thr Ala His Arg Ala

50 55 60

Gly Lys Lys His Leu Ser Ser Lys Leu Gly Gly Arg Arg Asp Gly Glu
65 70 75 80

Ala Thr Leu Glu Ile Ser Ala His His Ser Trp Cys Tyr Ala Phe Asn
85 90 95

Ser Val Ser Leu Ser Pro Gln Ala Leu Gln Leu Phe Tyr Gly Lys Lys
100 105 110

Gln Pro Gly Lys Glu Arg Lys Gln Asn Pro Lys His Gln Asn Glu Leu
115 120 125

Arg Arg Glu Glu Thr Lys Ala Glu Ala Pro Leu Leu Thr Gln Thr Arg
130 135 140

Leu Ile Thr Gln Ser Ala Leu His Arg Ala Pro His Tyr Asn Ser Cys
145 150 155 160

Cys Arg Arg Lys Tyr Arg Tyr Gly Thr Gly Lys Pro Glu Val
165 170

<210> 100
<211> 50
<212> PRT
<213> Homo sapien

<400> 100

Met Lys Tyr Pro Phe Ile Tyr Asn Tyr Phe Cys Leu Lys His Val Ser
1 5 10 15

Leu Tyr Ile Lys Asn Arg Tyr Phe Cys Phe His Phe Leu Ile Lys Phe
20 25 30

Cys Pro Tyr Phe Arg Ser Glu Lys Asn Gln Tyr Ser Asn Ile Lys Lys
35 40 45

Gln Glu
50

<210> 101
<211> 18
<212> PRT
<213> Homo sapien

<400> 101

Met Glu Glu Ile Tyr Leu Val Thr Gly Lys Leu Val Ile Gln Ala Leu
1 5 10 15

Glu Gly

<210> 102
<211> 34
<212> PRT
<213> Homo sapien

<400> 102

Met Ser Ser Gln Asn Arg Arg Cys Leu Gly Arg Asn Arg Gly Trp Cys
1 5 10 15

Leu Phe Ser Met Leu Ile Pro Tyr Pro Ser Asp Arg Ile Pro Phe Pro
20 25 30

Glu Val

<210> 103
<211> 40
<212> PRT
<213> Homo sapien

<400> 103

Met Asn Lys Gln Ile Tyr Cys Ser Ser Leu Lys Lys Phe Phe Phe Lys
1 5 10 15

Gln Ser His Ser Val Ala Gln Ala Gly Val Lys Gln Cys Asp Leu Ser
20 25 30

Ser Leu Gln Pro Pro Pro Pro Glu
35 40

<210> 104
<211> 990
<212> PRT
<213> Homo sapien

<400> 104

Met Ser Glu Glu Thr Arg Gln Ser Lys Leu Ala Ala Ala Lys Lys Lys
1 5 10 15

```
Leu Arg Glu Tyr Gln Gln Arg Asn Ser Pro Gly Val Pro Thr Gly Ala
            20                  25                  30

Lys Lys Lys Lys Lys Ile Lys Asn Gly Ser Asn Pro Glu Thr Thr Thr
        35                  40                  45

Ser Gly Gly Cys His Ser Pro Glu Asp Thr Pro Lys Asp Asn Ala Ala
    50                  55                  60

Thr Leu Gln Pro Ser Asp Asp Thr Val Leu Pro Gly Gly Val Pro Ser
65                  70                  75                  80

Pro Gly Ala Ser Leu Thr Ser Met Ala Ala Ser Gln Asn His Asp Ala
                85                  90                  95

Asp Asn Val Pro Asn Leu Met Asp Glu Thr Lys Thr Phe Ser Ser Thr
            100                 105                 110

Glu Ser Leu Arg Gln Leu Ser Gln Gln Leu Asn Gly Leu Val Cys Glu
        115                 120                 125

Ser Ala Thr Cys Val Asn Gly Glu Gly Pro Ala Ser Ser Ala Asn Leu
    130                 135                 140

Lys Asp Leu Glu Ser Arg Tyr Gln Gln Leu Ala Val Ala Leu Asp Ser
145                 150                 155                 160

Ser Tyr Val Thr Asn Lys Gln Leu Asn Ile Thr Ile Glu Lys Leu Lys
                165                 170                 175

Gln Gln Asn Gln Glu Ile Thr Asp Gln Leu Glu Glu Glu Lys Lys Glu
            180                 185                 190

Cys His Gln Lys Gln Gly Ala Leu Arg Glu Gln Leu Gln Val His Ile
        195                 200                 205

Gln Thr Ile Gly Ile Leu Val Ser Glu Lys Ala Glu Leu Gln Thr Ala
    210                 215                 220

Leu Ala His Thr Gln His Ala Ala Arg Gln Lys Glu Gly Glu Ser Glu
225                 230                 235                 240

Asp Leu Ala Ser Arg Leu Gln Tyr Ser Arg Arg Arg Val Gly Glu Leu
                245                 250                 255
```

Glu Arg Ala Leu Ser Ala Val Ser Thr Gln Gln Lys Lys Ala Asp Arg
260 265 270

Tyr Asn Lys Glu Leu Thr Lys Glu Arg Asp Ala Leu Arg Leu Glu Leu
275 280 285

Tyr Lys Asn Thr Gln Ser Asn Glu Asp Leu Lys Gln Glu Lys Ser Glu
290 295 300

Leu Glu Glu Lys Leu Arg Val Leu Val Thr Glu Lys Ala Gly Met Gln
305 310 315 320

Leu Asn Leu Glu Glu Leu Gln Lys Lys Leu Glu Met Thr Glu Leu Leu
325 330 335

Leu Gln Gln Phe Ser Ser Arg Cys Glu Ala Pro Asp Ala Asn Gln Gln
340 345 350

Leu Gln Gln Ala Met Glu Glu Arg Ala Gln Leu Glu Ala His Leu Gly
355 360 365

Gln Val Met Glu Ser Val Arg Gln Leu Gln Met Glu Arg Asp Lys Tyr
370 375 380

Ala Glu Asn Leu Lys Gly Glu Ser Ala Met Trp Arg Gln Arg Met Gln
385 390 395 400

Gln Met Ser Glu Gln Val His Thr Leu Arg Glu Glu Lys Glu Cys Ser
405 410 415

Met Ser Arg Val Gln Glu Leu Glu Thr Ser Leu Ala Glu Leu Arg Asn
420 425 430

Gln Met Ala Glu Pro Pro Pro Pro Glu Pro Pro Ala Gly Pro Ser Glu
435 440 445

Val Glu Gln Gln Leu Gln Ala Glu Ala Glu His Leu Arg Lys Glu Leu
450 455 460

Glu Gly Leu Ala Gly Gln Leu Gln Ala Gln Val Gln Asp Asn Glu Gly
465 470 475 480

Leu Ser Arg Leu Asn Arg Glu Gln Glu Glu Arg Leu Leu Glu Leu Glu
485 490 495

```
Arg Ala Ala Glu Leu Trp Gly Glu Gln Ala Glu Ala Arg Arg Gln Ile
            500                 505                 510

Leu Glu Thr Met Gln Asn Asp Arg Thr Thr Ile Ser Arg Ala Leu Ser
        515                 520                 525

Gln Asn Arg Glu Leu Lys Glu Gln Leu Ala Glu Leu Gln Ser Gly Phe
    530                 535                 540

Val Lys Leu Thr Asn Glu Asn Met Glu Ile Thr Ser Ala Leu Gln Ser
545                 550                 555                 560

Glu Gln His Val Lys Arg Glu Leu Gly Lys Lys Leu Gly Glu Leu Gln
                565                 570                 575

Glu Lys Leu Ser Glu Leu Lys Glu Thr Val Glu Leu Lys Ser Gln Glu
            580                 585                 590

Ala Gln Ser Leu Gln Gln Gln Arg Asp Gln Tyr Leu Gly His Leu Gln
        595                 600                 605

Gln Tyr Val Ala Ala Tyr Gln Gln Leu Thr Ser Glu Lys Glu Val Leu
    610                 615                 620

His Asn Gln Leu Leu Leu Gln Thr Gln Leu Val Asp Gln Leu Gln Gln
625                 630                 635                 640

Gln Glu Ala Gln Gly Lys Ala Val Ala Glu Met Ala Arg Gln Glu Leu
                645                 650                 655

Gln Glu Thr Gln Glu Arg Leu Glu Ala Ala Thr Gln Gln Asn Gln Gln
            660                 665                 670

Leu Arg Ala Gln Leu Ser Leu Met Ala His Pro Gly Glu Gly Asp Gly
        675                 680                 685

Leu Asp Arg Glu Glu Glu Glu Asp Glu Glu Glu Glu Glu Glu Glu Ala
    690                 695                 700

Val Ala Val Pro Gln Pro Met Pro Ser Ile Pro Glu Asp Leu Glu Ser
705                 710                 715                 720

Arg Glu Ala Met Val Ala Phe Phe Asn Ser Ala Val Ala Ser Ala Glu
```

725 730 735

Glu Glu Gln Ala Arg Leu Arg Gly Gln Leu Lys Glu Gln Arg Val Arg
740 745 750

Cys Arg Arg Leu Ala His Leu Leu Ala Ser Ala Gln Lys Glu Pro Glu
755 760 765

Ala Ala Ala Pro Ala Pro Gly Thr Gly Gly Asp Ser Val Cys Gly Glu
770 775 780

Thr His Arg Ala Leu Gln Gly Ala Met Glu Lys Leu Gln Ser Arg Phe
785 790 795 800

Met Glu Leu Met Gln Glu Lys Ala Asp Leu Lys Glu Arg Val Glu Glu
805 810 815

Leu Glu His Arg Cys Ile Gln Leu Ser Gly Glu Thr Asp Thr Ile Gly
820 825 830

Glu Tyr Ile Ala Leu Tyr Gln Ser Gln Arg Ala Val Leu Lys Glu Arg
835 840 845

His Arg Glu Lys Glu Glu Tyr Ile Ser Arg Leu Ala Gln Asp Lys Glu
850 855 860

Glu Met Lys Val Lys Leu Leu Glu Leu Gln Glu Leu Val Leu Arg Leu
865 870 875 880

Val Gly Asp Arg Asn Glu Trp His Gly Arg Phe Leu Ala Ala Ala Gln
885 890 895

Asn Pro Ala Asp Glu Pro Thr Ser Gly Ala Pro Ala Pro Gln Glu Leu
900 905 910

Gly Ala Ala Asn Gln Gln Gly Asp Leu Cys Glu Val Ser Leu Ala Gly
915 920 925

Ser Val Glu Pro Ala Gln Gly Glu Ala Arg Glu Gly Ser Pro Arg Asp
930 935 940

Asn Pro Thr Ala Gln Gln Ile Met Gln Leu Leu Arg Glu Met Gln Asn
945 950 955 960

Pro Arg Glu Arg Pro Gly Leu Gly Ser Asn Pro Cys Ile Pro Phe Phe
965 970 975

Tyr Arg Ala Asp Glu Asn Asp Glu Val Lys Ile Thr Val Ile
980 985 990

<210> 105
<211> 91
<212> PRT
<213> Homo sapien

<400> 105

Met Ala Pro Ala Val Pro Pro Arg Ala Ser Phe Phe Phe Phe Leu Leu
1 5 10 15

Phe Phe Phe Ile Phe Leu Leu Phe Lys Phe Tyr Trp Lys Phe Thr Asn
20 25 30

Val Leu Gln Thr Ser Val Lys His His Ile His Phe Thr Gly His Gly
35 40 45

Ser Gln Ala Ser Val Gln Asn Ser Leu Trp Gln Ser Pro His Gln Gly
50 55 60

Leu Leu Gln Thr Phe Leu Thr Asn Ser Leu Thr Leu Asn Thr Glu His
65 70 75 80

Arg Leu Trp Pro Ala Ser Pro Ser Gln Ala Leu
85 90

<210> 106
<211> 77
<212> PRT
<213> Homo sapien

<400> 106

Met Val Val Gly Gln Thr Pro His Thr Ser Val Leu Gln Lys His Ala
1 5 10 15

Phe Val Cys Glu Lys Pro Gln Pro Ala Pro Thr Ser Val Leu Gln Glu
20 25 30

Ala Trp Val Leu Gly Glu Glu Ala Pro Gly Gln Arg Pro Pro Ala Ser
35 40 45

Leu Gln Glu Ala Trp Gln Leu Tyr Val Arg Lys Pro Arg Pro Ala Pro

50 55 60

Thr Ser Val Pro Ala Gly Gln Ala Trp Thr Val Asn Gly
65 70 75

<210> 107
<211> 116
<212> PRT
<213> Homo sapien

<400> 107

Met Arg Gly Thr Pro Phe Leu Ser Cys Val Ala Cys Leu Val Cys Ala
1 5 10 15

Ser Thr Leu Leu Phe Leu Ser Leu Ser Ser Leu Lys Met Tyr Asn Lys
20 25 30

Ile Ser Phe Leu Ala Pro Arg Leu Ser Pro Pro Gln Asn Lys Lys Lys
35 40 45

Lys Lys Lys Lys Lys Asn Pro Phe Phe Phe Phe Phe Phe Phe Phe Leu
50 55 60

Phe Phe Phe Phe Phe Phe Phe Ala His Asn Lys Asn Leu Leu Gly Glu
65 70 75 80

Arg Trp Leu Met Gly Gly Lys Ile Trp Ile Gln Glu Ser Ser Ile Leu
85 90 95

Ala Leu Ala Leu Ser Pro Asn Pro Pro Ser Leu Pro Glu Pro Arg Gly
100 105 110

Val Ser Pro Cys
115

<210> 108
<211> 46
<212> PRT
<213> Homo sapien

<400> 108

Met Val Thr Leu Leu Phe Ser Glu Pro Leu Leu Arg Ala Ser Gln Asp
1 5 10 15

Ile Met Arg Thr Asp Asn Leu Pro Trp Ser Gln Arg Pro Ser Leu Pro
20 25 30

Leu Ala Arg Met Phe Arg Asp Arg Gln Arg Gly Gln Trp Trp
35 40 45

<210> 109
<211> 55
<212> PRT
<213> Homo sapien

<400> 109

Met Trp Glu Leu Thr Glu Gln Tyr His His Arg Val Asn Lys Leu Trp
1 5 10 15

Thr Lys Asp Lys Ala Gln Ser Phe Phe Phe Phe Phe Phe Phe Phe Phe
20 25 30

Arg Leu Ser Thr Leu Leu Ser Cys Pro Gln Ala Pro Arg Asn Ile Leu
35 40 45

Ser Pro His Leu Glu Thr Asp
50 55

<210> 110
<211> 876
<212> PRT
<213> Homo sapien

<400> 110

Ala Ser Ala Gly Ala Ala Gly Ser Leu Thr Arg Ser Pro Ser Ser Asp
1 5 10 15

Phe Gln Gly Ala Ser Val Glu Lys Lys Met Ala Gln Val Leu His Val
20 25 30

Pro Ala Pro Phe Pro Gly Thr Pro Gly Pro Ala Ser Pro Pro Ala Phe
35 40 45

Pro Ala Lys Asp Pro Asp Pro Pro Tyr Ser Val Glu Thr Pro Tyr Gly
50 55 60

Tyr Arg Leu Asp Leu Asp Phe Leu Lys Tyr Val Asp Asp Ile Glu Lys
65 70 75 80

Gly His Thr Leu Arg Arg Val Ala Val Gln Arg Arg Pro Arg Leu Ser
85 90 95

```
Ser Leu Pro Arg Gly Pro Gly Ser Trp Trp Thr Ser Thr Glu Ser Leu
            100                 105                 110

Cys Ser Asn Ala Ser Gly Asp Ser Arg His Ser Ala Tyr Ser Tyr Cys
        115                 120                 125

Gly Arg Gly Phe Tyr Pro Gln Tyr Gly Ala Leu Glu Thr Arg Gly Gly
    130                 135                 140

Phe Asn Pro Arg Val Glu Arg Thr Leu Leu Asp Ala Arg Arg Arg Leu
145                 150                 155                 160

Glu Asp Gln Ala Ala Thr Pro Thr Gly Leu Gly Ser Leu Thr Pro Ser
                165                 170                 175

Ala Ala Gly Ser Thr Ala Ser Leu Val Gly Val Gly Leu Pro Pro Pro
            180                 185                 190

Thr Pro Arg Ser Ser Gly Leu Ser Thr Pro Val Pro Pro Ser Ala Gly
        195                 200                 205

His Leu Ala His Val Arg Glu Gln Met Ala Gly Ala Leu Arg Lys Leu
    210                 215                 220

Arg Gln Leu Glu Glu Gln Val Lys Leu Ile Pro Val Leu Gln Val Lys
225                 230                 235                 240

Leu Ser Val Leu Gln Glu Glu Lys Arg Gln Leu Thr Val Gln Leu Lys
                245                 250                 255

Ser Gln Lys Phe Leu Gly His Pro Thr Ala Gly Arg Gly Arg Ser Glu
            260                 265                 270

Leu Cys Leu Asp Leu Pro Asp Pro Pro Glu Asp Pro Val Ala Leu Glu
        275                 280                 285

Thr Arg Ser Val Gly Thr Trp Val Arg Glu Arg Asp Leu Gly Met Pro
    290                 295                 300

Asp Gly Glu Ala Ala Leu Ala Ala Lys Val Ala Val Leu Glu Thr Gln
305                 310                 315                 320

Leu Lys Lys Ala Leu Gln Glu Leu Gln Ala Ala Gln Ala Arg Gln Ala
                325                 330                 335
```

```
Asp Pro Gln Pro Gln Ala Trp Pro Pro Pro Asp Ser Pro Val Arg Val
            340                 345                 350

Asp Thr Val Arg Val Val Glu Gly Pro Arg Glu Val Glu Val Val Ala
        355                 360                 365

Ser Thr Ala Ala Gly Ala Pro Ala Gln Arg Ala Gln Ser Leu Glu Pro
    370                 375                 380

Tyr Gly Thr Gly Leu Arg Ala Leu Ala Met Pro Gly Arg Pro Glu Ser
385                 390                 395                 400

Pro Pro Val Phe Arg Ser Gln Glu Val Val Glu Thr Met Cys Pro Val
                405                 410                 415

Pro Ala Ala Ala Thr Ser Asn Val His Met Val Lys Lys Ile Ser Ile
            420                 425                 430

Thr Glu Arg Ser Cys Asp Gly Ala Ala Gly Leu Pro Glu Val Pro Ala
        435                 440                 445

Glu Ser Ser Ser Ser Pro Pro Gly Ser Glu Val Ala Ser Leu Thr Gln
    450                 455                 460

Pro Glu Lys Ser Thr Gly Arg Val Pro Thr Gln Glu Pro Thr His Arg
465                 470                 475                 480

Glu Pro Thr Arg Gln Ala Ala Ser Gln Glu Ser Glu Glu Ala Gly Gly
                485                 490                 495

Thr Gly Gly Pro Pro Ala Gly Val Arg Ser Ile Met Lys Arg Lys Glu
            500                 505                 510

Glu Val Ala Asp Pro Thr Ala His Arg Arg Ser Leu Gln Phe Val Gly
        515                 520                 525

Val Asn Gly Gly Tyr Glu Ser Ser Ser Glu Asp Ser Ser Thr Ala Glu
    530                 535                 540

Asn Ile Ser Asp Asn Asp Ser Thr Glu Asn Glu Ala Pro Glu Pro Arg
545                 550                 555                 560

Glu Arg Val Pro Ser Val Ala Glu Ala Pro Gln Leu Arg Pro Ala Gly
```

565 570 575

Thr Ala Ala Ala Lys Thr Ser Arg Gln Glu Cys Gln Leu Ser Arg Glu
580 585 590

Ser Gln His Ile Pro Thr Ala Glu Gly Ala Ser Gly Ser Asn Thr Glu
595 600 605

Glu Glu Ile Arg Met Glu Leu Ser Pro Asp Leu Ile Ser Ala Cys Leu
610 615 620

Ala Leu Glu Lys Tyr Leu Asp Asn Pro Asn Ala Leu Thr Glu Arg Glu
625 630 635 640

Leu Lys Val Ala Tyr Thr Thr Val Leu Gln Glu Trp Leu Arg Leu Ala
645 650 655

Cys Arg Ser Asp Ala His Pro Glu Leu Val Arg Arg His Leu Val Thr
660 665 670

Phe Arg Ala Met Ser Ala Arg Leu Leu Asp Tyr Val Val Asn Ile Ala
675 680 685

Asp Ser Asn Gly Asn Thr Ala Leu His Tyr Ser Val Ser His Ala Asn
690 695 700

Phe Pro Val Val Gln Gln Leu Leu Asp Ser Gly Val Cys Lys Val Asp
705 710 715 720

Lys Gln Asn Arg Ala Gly Tyr Ser Pro Ile Met Leu Thr Ala Leu Ala
725 730 735

Thr Leu Lys Thr Gln Asp Asp Ile Glu Thr Val Leu Gln Leu Phe Arg
740 745 750

Leu Gly Asn Ile Asn Ala Lys Ala Ser Gln Ala Gly Gln Thr Ala Leu
755 760 765

Met Leu Ala Val Ser His Gly Arg Val Asp Val Val Lys Ala Leu Leu
770 775 780

Ala Cys Glu Ala Asp Val Asn Val Gln Asp Asp Asp Gly Ser Thr Ala
785 790 795 800

Leu Met Cys Ala Cys Glu His Gly His Lys Glu Ile Ala Gly Leu Leu
805 810 815

Leu Ala Val Pro Ser Cys Asp Ile Ser Leu Thr Asp Arg Asp Gly Ser
820 825 830

Thr Ala Leu Met Val Ala Leu Asp Ala Gly Gln Ser Glu Ile Ala Ser
835 840 845

Met Leu Tyr Ser Arg Met Asn Ile Lys Cys Ser Phe Ala Pro Met Ser
850 855 860

Asp Asp Glu Ser Pro Thr Ser Ser Ser Ala Glu Glu
865 870 875